

From the Department of Medicine, Solna
Karolinska Institutet, Stockholm, Sweden

EPIDEMIOLOGICAL ASPECTS OF SARCOIDOSIS: RISK FACTORS AND LONG-TERM CONSEQUENCES

Marios Rossides



**Karolinska
Institutet**

Stockholm 2021

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by Universitetsservice US-AB, 2021

© Marios Rossides, 2021

ISBN 978-91-8016-070-4

Cover illustration by Marios Rossides. Images used under license from Shutterstock.com

EPIDEMIOLOGICAL ASPECTS OF SARCOIDOSIS: RISK FACTORS AND LONG-TERM CONSEQUENCES

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Marios Rossides

The thesis will be defended in public on February 19, 2021, 13:00 at Marc Bygdeman room (J3:13), BioClinicum, Karolinska University Hospital, Solna, and online (via Zoom) at <https://bit.ly/MR-PhDThesisDefense>.

Principal Supervisor:

Docent Elizabeth V. Arkema
Karolinska Institutet
Department of Medicine, Solna
Clinical Epidemiology Division

Co-supervisors:

Professor Johan Grunewald
Karolinska Institutet
Department of Medicine, Solna
Respiratory Medicine Division

Professor Johan Askling

Karolinska Institutet
Department of Medicine, Solna
Clinical Epidemiology Division

Dr. Daniela Di Giuseppe

Karolinska Institutet
Department of Medicine, Solna
Clinical Epidemiology Division

Opponent:

Associate Professor Bo Baslund
University of Copenhagen
Department of Clinical Medicine

Examination Board:

Professor Dominique Valeyre
Université Sorbonne Paris Nord
Faculty of Health, Medicine and Human Biology
Hypoxia and Lung Research Unit

Professor Paul Dickman
Karolinska Institutet
Department of Medical Epidemiology
and Biostatistics

Docent Sara Öberg
Karolinska Institutet
Department of Medical Epidemiology
and Biostatistics

To my parents

«Σὰ βγεῖς στὸν πηγαμὸ γιὰ τὴν Ἰθάκη,
νὰ εὕχῃσαι νᾶναι μακρὺς ὁ δρόμος,
γεμάτος περιπέτειες, γεμάτος γνώσεις.

Τοὺς Λαιστρυγόνας καὶ τοὺς Κύκλωπας,
τὸν θυμωμένο Ποσειδῶνα μὴ φοβᾶσαι,
τέτοια στὸν δρόμο σου ποτὲ σου δὲν θὰ βρεῖς,
ἂν μὲν ᾗ σκέψις σου ὕψηλὴ, ἂν ἐκλεκτὴ
συγκίνησις τὸ πνεῦμα καὶ τὸ σῶμα σου ἀγγίζει.»

Κωνσταντῖνος Π. Καβάφης

POPULAR SCIENCE SUMMARY

Sarcoidosis – a benign inflammatory disease of infectious etiology?

Sarcoidosis causes fever, shortness of breath, cough, and tiredness in most patients. That is, symptoms like a common cold or the flu. Is sarcoidosis a benign disease like a common cold? And what causes the disease? These are the two broad topics dealt with in this doctoral thesis.

In sarcoidosis, non-cancerous masses known as granulomas form in various organs of the body. The lungs are primarily affected giving rise to general, flu-like symptoms from the respiratory tract. Granulomas are accumulations of several types of cells that try to surround and clear an unknown environmental insult. For several decades, microorganisms have been the main suspects for causing granuloma formation in individuals who have a genetic predisposition to the disease.

Genetic vulnerability and familial clustering of cases with sarcoidosis were examined in the first study in this thesis. Swedish national health and administrative databases were used to find individuals with and without sarcoidosis and compare the occurrence of the disease in their relatives. Relatives of sarcoidosis cases were more likely to have been diagnosed with the disease compared to relatives of unaffected individuals, which could be explained by shared genetic factors among related individuals. However, not all of the vulnerability to sarcoidosis could be explained by genetic similarities among family members. Environmental factors such as microorganisms may be etiologically linked with sarcoidosis occurrence.

The potential of infectious diseases increasing the risk of developing sarcoidosis was investigated in the second study in this thesis. Individuals with a history of infectious disease were slightly more likely to develop sarcoidosis in the future. Nevertheless, the etiologic role of infectious agents in sarcoidosis onset is still unclear. Our analyses suggested that subtle undetected inflammation likely caused by sarcoidosis in some patients might have increased their vulnerability to infectious diseases rather than the opposite.

After sarcoidosis diagnosis, one in two patients are expected to recover within a couple of years. The belief that sarcoidosis is benign in most patients and the limited availability of large and representative data sources did not allow for extensive investigation of key outcomes: early death, infectious disease, and heart failure. In the remaining four studies, we found that risks of early death, one or multiple hospitalizations for infectious disease, and heart failure were notably higher in sarcoidosis than in the general population, particularly within two years from sarcoidosis diagnosis. Interestingly, the highest risks for adverse outcomes were seen in patients with more severe sarcoidosis who needed treatment with immunosuppressants like prednisolone, methotrexate, or azathioprine. Patients with sarcoidosis who were started on methotrexate instead of azathioprine were less likely to be diagnosed with an infectious disease within six months from treatment initiation.

Overall, this doctoral thesis showed that familial disease is an important risk factor for sarcoidosis while infectious diseases might not play such a significant role in disease etiology as previously thought. Sarcoidosis is not a benign disease – a considerable proportion of patients with the disease are at high risks of debilitating complications such as infectious diseases and heart failure, which may reduce life expectancy in some. Increased vigilance for early diagnosis and administration of preventive measures is needed to tackle unfavorable outcomes in sarcoidosis.



POPULÄRVETENSKAPLIG SAMMANFATTNING

Sarkoidos – en godartad inflammatorisk sjukdom av infektiös etiologi?

Sarkoidos orsakar feber, andfåddhet, hosta och trötthet hos de flesta patienterna. Det vill säga symtom som en vanlig förkylning. Är sarkoidos en godartad sjukdom som en vanlig förkylning? Och vad orsakar sjukdomen? Dessa är de två breda ämnena som denna doktorsavhandling handlas om.

Vid sarkoidos bildas icke-cancermassor som kallas granulom i olika organ i kroppen. Lungorna påverkas främst och ger upphov till allmänna influensaliknande symtom från luftvägarna. Granulom är ansamlingar av inflammatoriska och andra typer av celler som försöker omge och rensa en okänd miljöförörlämpning. Under flera decennier har mikroorganismer varit de mest misstänkta för att orsaka granulombildning hos individer som har en genetisk benägenhet för sjukdomen.

Genetisk sårbarhet och familjära konstellationer av sarkoidosfall undersöktes i den första studien i denna avhandling. Svenska nationella hälso- och administrativa databaser användes för att hitta individer med och utan sarkoidos och jämföra förekomsten av sjukdomen hos sina släktingar. Släktingar till sarkoidosfall var mer benägna att ha sjukdomen jämfört med släktingar till opåverkade individer. Detta kan förklaras av genetiska faktorer som delas mellan släktingar. Trots det förklarades inte all risk för att insjukna med sarkoidos av genetiska faktorer. Miljöfaktorer som mikroorganismer kan dessutom vara etiologiskt kopplade till förekomsten av sarkoidos.

Om infektionssjukdomar kan öka risken för att insjukna med sarkoidos undersöktes i den andra studien i denna avhandling. Individer med infektionssjukdomar i sin anamnes var något mer benägna att diagnostiseras med sarkoidos i framtiden. Ändå är rollen som infektiösa medel spelar i sarkoidos etiologi fortfarande oklar. Våra analyser föreslog att subtilt upptäckt sarkoidos-relaterade inflammation hos vissa individer kan ha ökat sin sårbarhet för infektion snarare än tvärtom.

En av två patienter förväntas återhämta sig inom ett par år efter sarkoidosdiagnos. Tron att sjukdomen är godartad för de flesta patienter och den begränsade tillgången till stora och representativa datakällor möjliggjorde inte omfattande undersökningar av viktiga kliniska fall: tidig död, infektionssjukdomar och hjärtsvikt. I de återstående fyra studierna fann vi att riskerna för tidig död, en eller flera sjukhusvistelser för infektionssjukdom och hjärtsvikt var högre vid sarkoidos än i den allmänna befolkningen särskilt inom två år från sarkoidosdiagnos. I synnerhet upptäcktes de högsta riskerna för komplikationer hos patienter med allvarligare sarkoidos som behövde behandling med immunsuppressiva läkemedel som prednisolon, metotrexat eller azatioprin. Patienter med sarkoidos som behandlades med metotrexat istället för azatioprin var dock mindre benägna att diagnostiseras med en infektionssjukdom inom sex månader efter terapins start.

Sammanfattningsvis visade denna doktorsavhandling att familjesjukdom är en viktig riskfaktor för sarkoidos medan smittsamma sjukdomar kanske inte spelar en signifikant roll i sjukdomsetiologi som tidigare trott. Sarkoidos är inte en godartad sjukdom – en stor del av sarkoidospatienter löper risk för allvarliga komplikationer som infektionssjukdomar och hjärtsvikt vilket kan minska livslängden hos vissa. Ökad vaksamhet för tidig diagnos och tillhandahållande av förebyggande åtgärder behövs för att minska de höga riskerna för ogynnsamma komplikationer vid sarkoidos.

ΠΕΡΙΛΗΨΗ

Σαρκοείδωση, μια καλοήθους φλεγμονώδους νόσος λοιμώδους αιτιολογίας;

Στους περισσότερους ασθενείς η σαρκοείδωση εμφανίζεται με πυρετό, δύσπνοια, βήχα και κόπωση. Παρόμοια δηλαδή συμπτώματα με ένα κοινό κρυολόγημα ή γρίπη. Είναι η σαρκοείδωση μια καλοήθους ασθένεια σαν ένα κοινό κρυολόγημα; Και τι προκαλεί τη νόσο αυτή; Αυτά είναι τα δύο ευρέα θέματα με τα οποία ασχολείται αυτή η διδακτορική διατριβή.

Μη καρκινικές μάζες γνωστές ως κοκκιώματα σχηματίζονται σε διάφορα όργανα του σώματος σε ασθενείς με σαρκοείδωση. Οι πνεύμονες επηρεάζονται στην πλειονότητα των ασθενών προκαλώντας γενικά συμπτώματα από το αναπνευστικό σύστημα τα οποία προσομοιάζουν με συμπτώματα γρίπης. Τα κοκκιώματα αποτελούν συσσωρεύσεις κυττάρων στην προσπάθεια του οργανισμού να περιβάλει και να εξαλείψει ένα ή περισσότερα άγνωστα μέχρι σήμερα παθογόνα. Εδώ και δεκαετίες, μικροοργανισμοί όπως βακτηρίδια αποτελούν τους κύριους υπόπτους για την πρόκληση σχηματισμού κοκκιωμάτων σε άτομα με γενετική προδιάθεση για την ασθένεια.

Η γενετική ευπάθεια και η συσσώρευση οικογενών περιπτώσεων σαρκοείδωσης εξετάστηκαν στην πρώτη μελέτη που περιλαμβάνεται σε αυτή τη διατριβή. Χρησιμοποιήθηκαν βάσεις δεδομένων με ιατρικά και δημογραφικά στοιχεία από τη Σουηδία για την εύρεση ατόμων με και χωρίς σαρκοείδωση και σύγκριση της εμφάνισης της νόσου στους συγγενείς των. Οι συγγενείς των ασθενών με σαρκοείδωση είχαν περισσότερες πιθανότητες να πάσχουν από τη νόσο σε σύγκριση με συγγενείς ατόμων χωρίς σαρκοείδωση. Αυτό θα μπορούσε να εξηγηθεί από την ύπαρξη γενετικής προδιάθεσης για τη νόσο μεταξύ συγγενικών ατόμων. Ωστόσο, στη μελέτη αυτή η προδιάθεση για σαρκοείδωση δεν μπόρεσε να εξηγηθεί μόνο από τη γενετική ομοιότητα μεταξύ ασθενών και των συγγενών τους. Ως εκ τούτου, περιβαλλοντικοί παράγοντες όπως μικροοργανισμοί μπορεί να συνδέονται αιτιολογικά με την εμφάνιση σαρκοείδωσης.

Η πιθανότητα λοιμώδεις ασθένειες να αυξάνουν τον κίνδυνο εμφάνισης σαρκοείδωσης διερευνήθηκε στη δεύτερη μελέτη σε αυτή τη διατριβή. Άτομα με ιστορικό λοιμωδών ασθενειών είχαν ελαφρώς περισσότερες πιθανότητες να αναπτύξουν σαρκοείδωση στο μέλλον. Παρόλα αυτά, ο ρόλος των λοιμογόνων παραγόντων στην πρόκληση σαρκοείδωσης είναι ακόμα ασαφής. Η ευπάθεια ορισμένων ασθενών σε λοίμωξη από μικροοργανισμούς πιθανολογείται να είναι αποτέλεσμα μη ανιχνεύσιμης φλεγμονής προκαλούμενης από τη σαρκοείδωση και όχι το αντίθετο.

Ένας στους δύο ασθενείς με σαρκοείδωση αναμένεται να αναρρώσει μέσα σε δύο χρόνια από τη διάγνωση της νόσου. Η πεποίθηση ότι για τους περισσότερους ασθενείς η σαρκοείδωση είναι καλοήθους όπως επίσης και η περιορισμένη διαθεσιμότητα μεγάλων και αντιπροσωπευτικών πηγών δεδομένων δεν επέτρεψαν μέχρι σήμερα την εκτεταμένη διερεύνηση της έκβασης της νόσου. Οι υπόλοιπες τέσσερις μελέτες οι οποίες περιλαμβάνονται στη διατριβή αυτή έδειξαν ότι ο κίνδυνος για πρόωπο θάνατο, μία ή πολλαπλές νοσηλείες για

λοιμώδεις ασθένειες και για καρδιακή ανεπάρκεια ήταν σημαντικά υψηλότερος στους ασθενείς με σαρκοείδωση από ότι στο γενικό πληθυσμό, ιδιαίτερα εντός δύο ετών από τη διάγνωση της νόσου. Μεγαλύτερο ρίσκο για πρόωρο θάνατο, συννοσηρότητα από λοιμώδεις ασθένειες και καρδιακή ανεπάρκεια παρατηρήθηκαν σε ασθενείς με σοβαρής μορφής σαρκοείδωση οι οποίοι χρειάστηκαν θεραπεία με ανοσοκατασταλτικά φάρμακα όπως κορτικοειδή, μεθοτρεξάτη ή αζαθειοπρίνη. Ωστόσο, ασθενείς οι οποίοι λάμβαναν μεθοτρεξάτη αντί της αζαθειοπρίνης είχαν λιγότερες πιθανότητες να διαγνωστούν με κάποια λοιμώδη νόσο εντός έξι μηνών από την έναρξη λήψης της θεραπείας.

Συμπερασματικά, η διδακτορική διατριβή αυτή έδειξε ότι η οικογενής προδιάθεση για σαρκοείδωση είναι ένας σημαντικός παράγοντας κινδύνου για τη νόσο. Αντίθετα, οι λοιμώδεις νόσοι μπορεί να μη διαδραματίζουν τόσο σημαντικό ρόλο στην αιτιολογία της νόσου όπως πιστευόταν μέχρι σήμερα. Η σαρκοείδωση δεν είναι καλοήθης ασθένεια όπως κάποιοι πιστεύουν. Ορισμένοι ασθενείς διατρέχουν κίνδυνο εμφάνισης ανεπιθύμητων επιπλοκών όπως λοιμωδών ασθενειών και καρδιακής ανεπάρκειας οι οποίες ενδέχεται να μειώσουν το προσδόκιμο ζωής σε σχετικά μεγάλη αναλογία ασθενών με τη νόσο. Για το λόγο αυτό απαιτείται αυξημένη επαγρύπνηση για έγκαιρη διάγνωση και χορήγηση προληπτικών μέτρων σε αυτούς τους ασθενείς για την αντιμετώπιση δυσμενών επιπλοκών οι οποίες σχετίζονται με τη σαρκοείδωση.

公共科学の概要

サルコイドーシス - 感染性の良性炎症性疾患？

サルコイドーシスは、ほとんどの患者に発熱、息切れ、咳、倦怠感を引き起こす。つまり、風邪やインフルエンザのような症状である。サルコイドーシスは風邪のような良性の病気なのか？そして、何が原因で発症するのか？この2つが本論文のテーマである。

サルコイドーシスでは、肉芽腫と呼ばれる非がん性の腫瘍が体の様々な器官に形成される。肺が主に侵され、気道からの全身的なインフルエンザ様症状が生じる。肉芽腫は、未知の環境障害を取り囲んで除去しようとする炎症性の細胞やその他の細胞の集まりである。数十年前から、微生物がこの病気の遺伝的素因を持つ人の肉芽腫形成を引き起こす主な原因であると考えられてきた。

本論文の最初の研究では、サルコイドーシス患者の遺伝的脆弱性と家族集積性について検討した。スウェーデンの国民健康データベースと行政データベースを用いて、サルコイドーシスの罹患者と罹患していない者を検索し、その親族におけるサルコイドーシスの発生状況を比較した。その結果、サルコイドーシス患者の親族は、罹患していない親族に比べてサルコイドーシスと診断された可能性が高く、これは親族間で遺伝的要因が共有されていることによって説明できる可能性があった。しかし、サルコイドーシスに対する脆弱性のすべてが家族間の遺伝的類似性によって説明できるわけではなかった。微生物などの環境因子がサルコイドーシス発症と病的に関連している可能性がある。

感染症がサルコイドーシスを発症するリスクを高める可能性については、本論文の2番目の研究で調査した。感染症の既往歴のある個人は、将来サルコイドーシスを発症する可能性がわずかに高かった。それにもかかわらず、サルコイドーシス発症における感染症の病因的役割はまだ明らかにされていない。我々の解析では、サルコイドーシスによって引き起こされたとされる微妙な炎症が、むしろ反対に、一部の患者では感染に対する脆弱性を高めている可能性が示唆された。

サルコイドーシスの診断後、2人に1人の患者は2年以内に回復すると予想されている。ほとんどの患者ではサルコイドーシスは良性であると考えられていることや、大規模で代表的なデータソースが限られていることから、主要な転帰である早期死亡、感染症、心不全についての広範な調査を行うことができなかった。残りの4件の研究では、早期死亡、感染症による1回または複数回の入院、心不全のリスクが一般集団よりもサルコイドーシスで顕著に高く、特にサルコイドーシスの診断から2年以内であることが明らかになった。特に、有害転帰のリスクが最も高かったのは、プレドニゾロン、メトトレキサート、アザチオプリンなどの免疫抑制剤による

治療を必要とする、より重症のサルコイドーシス患者であった。しかし、アザチオプリンの代わりにメトトレキサートで治療を開始した患者は、治療開始から6ヵ月以内に感染症と診断される可能性が低かった。

全体的に、この博士論文は、家族性疾患がサルコイドーシスの重要な危険因子である一方で、感染症はこれまで考えられていたような病因には重要な役割を果たしていない可能性があることを示しています。サルコイドーシスは良性の病気ではなく、かなりの割合の患者が感染症や心不全などの衰弱性合併症のリスクが高く、一部の患者では余命が短くなる可能性がある。サルコイドーシスの好ましくない結果に取り組むためには、早期診断と予防策の投与に対する警戒心を高める必要がある。

ABSTRACT

Sarcoidosis is a systemic inflammatory disease of unknown etiology in which granulomatous lesions form mostly in the lungs and the lymphatic system of patients. Although more than a century has passed since sarcoidosis was first described, our understanding of its etiology and clinical course is limited. That is because epidemiological studies on large and representative patient cohorts have been lacking. The scope of this thesis was to examine aspects of sarcoidosis epidemiology using a linkage of large, nationwide health and administrative databases from Sweden complemented by clinical data. Six individual studies are included in this thesis; the first two dealt with risk factors for sarcoidosis, namely familial and infectious disease, and the rest with long-term debilitating patient outcomes: mortality, infection, and heart failure.

In Study I, a case-control-family study, we estimated familial relative risks and the heritability of sarcoidosis. We found that having first-degree relatives with sarcoidosis increased the risk of being diagnosed with the disease by more than threefold. 39% of the susceptibility to sarcoidosis in the Swedish population was estimated to be attributable to additive genetic effects; the rest was due to non-shared (among siblings) environmental factors.

Study II was a case-control study in which we estimated relative risks of sarcoidosis associated with having a history of infectious disease diagnoses. We showed that infectious diseases (commonly upper respiratory and genitourinary) diagnosed before sarcoidosis diagnosis were associated with a small increased risk of sarcoidosis in the future, a relative risk that did not vary markedly by latency period between infectious disease ascertainment and sarcoidosis diagnosis. These small relative risks could be easily explained away in analyses designed to test the robustness of these associations in the presence of reverse causation bias.

In Study III, a cohort study, we followed individuals with sarcoidosis and general population comparators for all-cause death. We showed that there was an overall 61% increased risk of death associated with sarcoidosis. Stratification by treatment status around the time of sarcoidosis diagnosis approximating disease severity revealed a 2.3-fold higher risk of all-cause mortality compared to the general population in those treated while no risk increase was observed for untreated patients with sarcoidosis.

Similarly, in Study IV, we followed individuals for a first or recurrent serious (hospitalized) infections. We observed a 1.8-fold higher risk of serious infection in sarcoidosis compared to the general population, which was even higher during the first two years since diagnosis and in individuals who were treated with an immunosuppressant around sarcoidosis diagnosis likely due to more severe or progressive disease at the time.

In Study V, a target trial emulation, we compared six-month risks of infectious disease in initiators of methotrexate compared to azathioprine, two second line treatments for sarcoidosis. Six months after treatment initiation, a 43% lower risk of infectious disease was observed in the methotrexate compared to the azathioprine group.

Study VI was a cohort study in which we examined the relative risk of heart failure and its predictors in sarcoidosis. We found a 2.4-fold increased relative risk of heart failure associated with sarcoidosis that was higher during the first two years since sarcoidosis diagnosis and in individuals without a history of ischemic heart disease. Diabetes, atrial fibrillation, and other arrhythmias were the strongest clinical predictors of heart failure diagnosis in sarcoidosis.

Overall, findings from studies on risk factors in this thesis suggest that familial disease and genetics are important in sarcoidosis, albeit a larger contribution to the etiology of sarcoidosis is likely due to environmental factors. Among environmental factors, clinically identifiable infectious diseases are unlikely to be strong risk factors for sarcoidosis diagnosis. Future molecular and epidemiological studies on environmental triggers of sarcoid inflammation and disease should consider the issue of reverse causality owing to long preclinical disease in some patients. Studies on long-term patient outcomes in this thesis showed that sarcoidosis is not a ‘benign’ disease. Therefore, our quest to identify effective interventions and groups of patients to target should continue. If applied early, these measures can help alleviate some of the risks related to infection and heart failure, and improve life expectancy, especially in patients with severe or chronic disease.

Keywords: sarcoidosis; risk factors; mortality; infections; heart failure; registries; Sweden

LIST OF SCIENTIFIC PAPERS

- I. Rossides M, Grunewald J, Eklund A, Kullberg S, Di Giuseppe D, Askling J, Arkema EV. **Familial aggregation and heritability of sarcoidosis: a Swedish nested case-control study.** *European Respiratory Journal*, 2018; 52:1800385. <https://doi.org/10.1183/13993003.00385-2018>
- II. Rossides M, Kullberg S, Askling J, Eklund A, Grunewald J, Di Giuseppe D, Arkema EV. **Are infectious diseases risk factors for sarcoidosis or a result of reverse causation? Findings from a population-based nested case-control study.** *European Journal of Epidemiology*, 2020; 35:1087–97. <https://doi.org/10.1007/s10654-020-00611-w>
- III. Rossides M, Kullberg S, Askling J, Eklund A, Grunewald J, Arkema EV. **Sarcoidosis mortality in Sweden: a population-based cohort study.** *European Respiratory Journal*, 2018; 51:1701815. <https://doi.org/10.1183/13993003.01815-2017>
- IV. Rossides M, Kullberg S, Eklund A, Di Giuseppe D, Grunewald J, Askling J, Arkema EV. **Risk of first and recurrent serious infection in sarcoidosis: a Swedish register-based cohort study.** *European Respiratory Journal*, 2020; 56:2000767. <https://doi.org/10.1183/13993003.00767-2020>
- V. Rossides M, Kullberg S, Di Giuseppe D, Eklund A, Grunewald J, Askling J, Arkema EV. **Infection risk in sarcoidosis patients treated with methotrexate compared to azathioprine: A retrospective ‘target trial’ emulated with Swedish real-world data.** *Respirology*, 2021; [In press]. <https://doi.org/10.1111/resp.14001>
- VI. Rossides M, Kullberg S, Grunewald J, Eklund A, Di Giuseppe D, Askling J, Arkema EV. **Risk and predictors of heart failure in sarcoidosis: a population-based cohort study from Sweden.** [Manuscript].

SCIENTIFIC PAPERS NOT INCLUDED IN THE THESIS

- VII. Rossides M, Kullberg S, Eklund A, Grunewald J, Arkema EV. **Sarcoidosis diagnosis and treatment in Sweden: a register-based assessment of variations by region and calendar period.** *Respiratory Medicine*, 2020; 161:105846. <https://doi.org/10.1016/j.rmed.2019.105846>
- VIII. Köcher L, Rossides M, Remaeus K, Grunewald J, Eklund A, Kullberg S, Arkema EV. **Maternal and infant outcomes in sarcoidosis pregnancy: a Swedish population-based cohort study of first births.** *Respiratory Research*, 2020; 21:225. <https://doi.org/10.1186/s12931-020-01493-y>
- IX. Ceder S, Rossides M, Kullberg S, Eklund A, Grunewald J, Arkema EV. **Positive predictive value of sarcoidosis identified in an administrative healthcare registry: a validation study.** *Epidemiology*, 2021; [In press].
- X. Rossides M, Kullberg S, Grunewald J, Eklund A, Di Giuseppe D, Askling J, Arkema EV. **Risk of acute myocardial infarction in sarcoidosis: a population-based cohort study from Sweden.** [Manuscript].

CONTENTS

1	INTRODUCTION	1
1.1	Sarcoidosis: A heterogenous inflammatory disease	1
1.2	Sarcoidosis pathophysiology and immunopathogenesis	2
1.3	Löfgren's syndrome and cardiac sarcoidosis.....	3
1.4	Sarcoidosis diagnosis	5
1.5	Sarcoidosis treatment	8
1.6	Clinical course of sarcoidosis	9
2	LITERATURE REVIEW	11
2.1	Epidemiology of sarcoidosis: An introduction.....	11
2.2	Sarcoidosis occurrence.....	11
2.3	Risk factors for sarcoidosis	14
2.3.1	Genetics of sarcoidosis	14
2.3.2	Environmental exposures predisposing to sarcoidosis.....	16
2.4	Outcomes of patients with sarcoidosis.....	19
2.4.1	Mortality.....	19
2.4.2	Serious infections	20
2.4.3	Cardiovascular disease and heart failure	21
2.4.4	Other somatic and mental health outcomes.....	22
2.5	Knowledge gaps in sarcoidosis epidemiology.....	22
3	RESEARCH AIMS.....	23
4	METHODS	25
4.1	Overview of epidemiological methods	25
4.2	Overview of statistical methods	29
4.3	Healthcare in Sweden.....	30
4.4	Data sources	32
4.5	Study design and statistical analyses	33
4.5.1	Sarcoidosis cases and comparators.....	34
4.5.2	Study I: Familial aggregation of sarcoidosis	35
4.5.3	Study II: Infectious diseases as risk factors for sarcoidosis	36
4.5.4	Study III: Mortality in sarcoidosis	38
4.5.5	Study IV: Risk of serious infection in sarcoidosis	39
4.5.6	Study V: Infection in methotrexate versus azathioprine for sarcoidosis	40
4.5.7	Study VI: Risk and predictors of heart failure in sarcoidosis	43
4.6	Ethical perspectives.....	45
5	RESULTS	47
5.1	The sarcoidosis group at diagnosis	47
5.2	Main findings by individual study	48
5.2.1	Study I: Familial aggregation of sarcoidosis	48
5.2.2	Study II: Infectious diseases as risk factors for sarcoidosis	50
5.2.3	Study III: Mortality in sarcoidosis	52

5.2.4	Study IV: Risk of serious infection in sarcoidosis	54
5.2.5	Study V: Infection in methotrexate versus azathioprine for sarcoidosis.....	56
5.2.6	Study VI: Risk and predictors of heart failure in sarcoidosis	57
6	DISCUSSION.....	61
6.1	The findings of this thesis in context	61
6.1.1	Risk factors for sarcoidosis.....	61
6.1.2	Long-term consequences of sarcoidosis.....	65
6.1.3	Transportability of findings	69
6.2	Methodological considerations.....	69
6.2.1	Registers in clinical epidemiologic research: friend (and foe?).....	69
6.2.2	Limitations of concepts and methods	71
7	CONCLUSIONS	75
8	POINTS OF PERSPECTIVE.....	77
9	ACKNOWLEDGEMENTS	83
10	REFERENCES	87

LIST OF ABBREVIATIONS

ACCESS	A Case-Control Epidemiologic Study of Sarcoidosis
CI	Confidence interval
DNA	Deoxyribonucleic acid
HLA	Human leukocyte antigen
ICD	International Classification of Diseases
NPR	National Patient Register
PDR	Prescribed Drug Register

1 INTRODUCTION

1.1 SARCOIDOSIS: A HETEROGENOUS INFLAMMATORY DISEASE

Sarcoidosis is a relatively rare granulomatous disease of unknown etiology. Sarcoid granulomas are organized round masses consisting of inflammatory cells that can develop in any organ or system of the human body [1]. However, the disease is primarily pulmonary as the lungs and lymph nodes of the thoracic cavity are affected in more than 90% of patients [2]. Extrapulmonary localizations can co-exist with pulmonary disease and include, among others, the skin, the eyes (ocular sarcoidosis), the heart (cardiac sarcoidosis), and the central nervous system (neurosarcoidosis) [2].

Sarcoidosis is extremely heterogeneous in terms of onset. Disease phenotypes range from completely asymptomatic with pulmonary alterations found by chance on routine chest radiographs to subacute and acute clinical pictures [3]. In about 50–70% of individuals, the onset of sarcoidosis is subacute with general symptoms (e.g. low-grade fever, extreme fatigue, night sweats, etc.) combined with symptoms from the affected organ (e.g., persistent cough, dyspnea or skin alterations) [3]. In a subset of patients (20–40%), the onset is abrupt with fever, erythema nodosum (skin lesions), with or without arthralgias affecting mostly the ankles. In these patients, bilateral hilar lymphadenopathy found on chest imaging completes the classical clinical picture of Löfgren's syndrome [4]. About 10% of individuals with sarcoidosis are asymptomatic and may be identified incidentally during routine chest imaging [5,6]. Asymptomatic sarcoidosis was likely more prevalent in older patient cohorts when mass radiography as a measure to control tuberculosis transmission in the population was more common.

The first description of sarcoid lesions dates back to the late 1800's when Jonathan Hutchinson (1828–1913), London physician and Professor of Medicine, described a patient with what is now believed to be cutaneous sarcoidosis [7,8]. 'Mortimer's malady' became known in 1898 after Dr. Hutchinson's other patient with cutaneous sarcoidosis resembling lupus pernio (a rare cutaneous form of sarcoidosis), but due to her disappearing, Ms. Mortimer's skin lesions were not histologically examined [7]. A year later, in 1889, Drs. Besnier (1831–1909) and Tenneson described the basic histological features of the sarcoid granuloma in skin biopsies obtained from two of their patients with lupus pernio [7]. In the late 1890's, Dr. Ernest Boeck (1845–1917) presented 24 cases in Oslo with lesions in various organs and body systems that resembled sarcomas, a cancer of connective tissue origin [7,8]. 'Mortimer's malady' became 'sarcoidosis' and the skin disease was understood to be a systemic disease instead.

Five decades later, it was the seminal work of Dr. Sven Löfgren (1910–1978) in Stockholm that distinguished sarcoidosis from tuberculosis (a granulomatous infectious disease with lung predominance) and highlighted the higher than previously perceived prevalence of sarcoidosis and its likely good prognosis [7]. He was also the first to link erythema nodosum and bilateral hilar lymphadenopathy (in chest X-rays), an acute form of sarcoidosis with favorable prognosis, and to describe hypercalcemia and renal involvement in sarcoidosis [9].

1.2 SARCOIDOSIS PATHOPHYSIOLOGY AND IMMUNOPATHOGENESIS

Histologically, the sarcoid granuloma is usually well-developed and is characterized by a central collection of epithelioid cells and Langhans or foreign-body type giant cells with multiple nuclei [10,11]. Lymphocytes surround the periphery and central necrosis is minimal (Figure 1). Schaumann bodies, asteroid bodies, and birefringent crystalline particles (e.g., calcium oxalate salt accumulations) may also be present in the sarcoid granuloma [10]. Granulomas are, however, not unique to sarcoidosis thus careful histological characterization and rigorous differential diagnosis are essential for correctly diagnosing the disease [11].

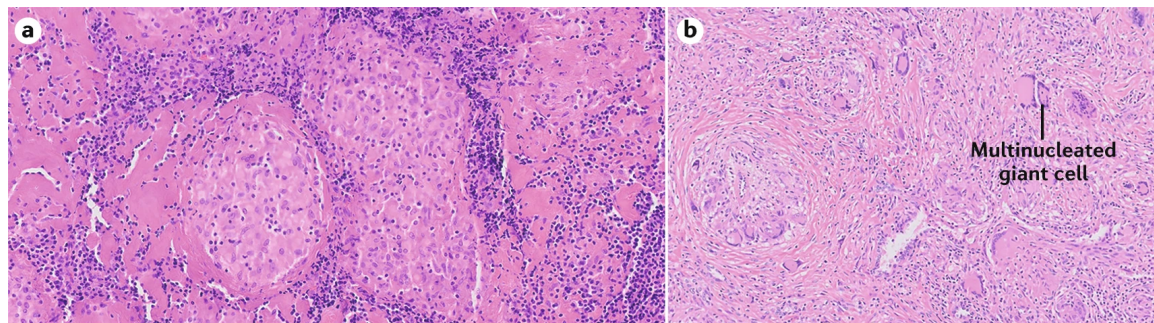


Figure 1 | The sarcoid granuloma.

- a) Biopsy specimen taken from an enlarged mediastinal lymph node showing non-necrotizing granulomas in a patient with radiographic type I sarcoidosis on a chest radiograph. Magnification 200×.
- b) Specimen from a consolidated mass in the lung of a patient with pulmonary sarcoidosis showing non-necrotizing granulomas with multinucleated giant cells. Magnification 100×. Biopsy samples in both panels were stained with hematoxylin and eosin. Images courtesy of C. A. Seldenrijk, St Antonius Hospital, The Netherlands. (Reprinted by permission from Springer Nature: *Nature Reviews Disease Primers*, “Sarcoidosis” by Grunewald J, Grutters JC, Arkema EV, et al. Copyright, 2019.)

The sequelae of events behind granuloma formation in sarcoidosis has not been completely elucidated due to the lack of a reliable mouse model for the disease [12]. The role of the draining lymph node in the lung, and likely in other organs, is perceived to be cardinal in the process of sarcoid granuloma formation [12]. Antigen-presenting cells such as interstitial dendritic cells in the lung appear to migrate to the mediastinal lymph nodes and through an HLA-mediated process, they present an unknown antigen to naïve T cells [12–14]. In turn, these T cells clonally expand to T helper 1 and T helper 17 cells that later migrate and orchestrate the inflammatory process resulting in granuloma formation [12].

Meanwhile other antigen presenting cells such as alveolar macrophages activate and produce inflammatory chemokines (e.g., tumor necrosis factor α and interferon γ) when in contact with the putative antigen. These chemokines enable and enhance the migration of T helper 1 and 17 cells, regulatory T cells, and to a lesser extent, B cells to the zone of inflammation [12,15]. The continuous and unmitigated presumably by regulatory T cells production of chemokines such as interleukin 6, 12, and 23, and transforming growth factor β in the area of inflammation, enhances the influx of inflammatory cells leading to granuloma formation [12].

The primary involvement of the lung and its lymphatic drainage system in sarcoidosis has since long raised the notion that an aerosolized agent is responsible for triggering inflammation and granuloma formation in genetically susceptible individuals [1]. Throughout the years, many exogenous agents, both organic (e.g., bacteria and viruses) and inorganic (e.g., metals), were investigated, but no single one appears to be wholly responsible for the abnormal inflammatory process observed in sarcoidosis [14]. Exogenous exposures as risk factors for sarcoidosis are extensively discussed later in section 2.3.2 (page 16).

Accumulating data during the last two decades suggests that several endogenous (self-) antigens (e.g., vimentin and serum amyloid A) may be potential triggers of sarcoid inflammation. This could shift our understanding of sarcoidosis as being an ‘infectious’ disease to sarcoidosis being an inflammatory disease with autoimmune features [14,15]. The quest to identify the etiologic agent or agents that initiate and maintain the inflammatory process in sarcoidosis is ongoing.

1.3 LÖFGREN'S SYNDROME AND CARDIAC SARCOIDOSIS

Löfgren's syndrome and cardiac sarcoidosis are two phenotypes of sarcoidosis that were investigated in individual studies in this thesis and therefore merit a brief description. Löfgren's syndrome refers to the combination of bilateral mediastinal lymphadenopathy in chest imaging with abrupt onset of fever, erythema nodosum (i.e. erythematous and tender nodular lesions commonly located on the shins; Figure 2) in female patients, and ankle arthralgias owing to periarticular inflammation that is more frequent in male patients [4]. In more than 80% of patients with Löfgren's sarcoidosis, symptoms and signs of disease are expected to resolve within two years from diagnosis with low risk of reemergence (<6%) [16]. Emerging evidence from genetic and immunologic studies illustrate that Löfgren's syndrome is distinctly different from other forms of sarcoidosis that have a more insidious onset and chronic course of disease [4].



Figure 2 | Erythema nodosum in a patient with Löfgren's syndrome.

(Reprinted by permission from Georg Thieme Verlag KG: *Seminars in Respiratory and Critical Care Medicine*, “Löfgren's Syndrome: Diagnosis, Management, and Disease Pathogenesis” by Karakaya B, Kaiser Y, van Moorsel CHM, et al. Copyright, 2017.)

Cardiac sarcoidosis is a severe and potentially lethal presentation of sarcoidosis in which sarcoid granulomatous inflammation infiltrates predominantly the myocardium [17]. Lesions are frequently found in the left ventricular wall especially at the basal level and septal segments as well as the right ventricular free wall [18] (Figure 3). Clinically overt disease characterized by ventricular arrhythmias and high-grade heart blocks is observed in about 5% of cases with sarcoidosis resulting in heart failure and/or sudden cardiac death [19–21]. It is believed, however, that cardiac sarcoidosis affects more than 30% of individuals with the disease as demonstrated by autopsy studies [19]. In recent years, modern imaging techniques such as positron emission tomography combined with computed tomography and cardiac magnetic resonance imaging have improved the diagnosis of covert cardiac sarcoidosis [17,22–24]. However, current expert consensus guidelines from the Heart Rhythm Society do not recommend cardiac imaging as a screening tool for patients without symptoms or electrocardiographic or echocardiographic findings that signify potential involvement of the heart [22].

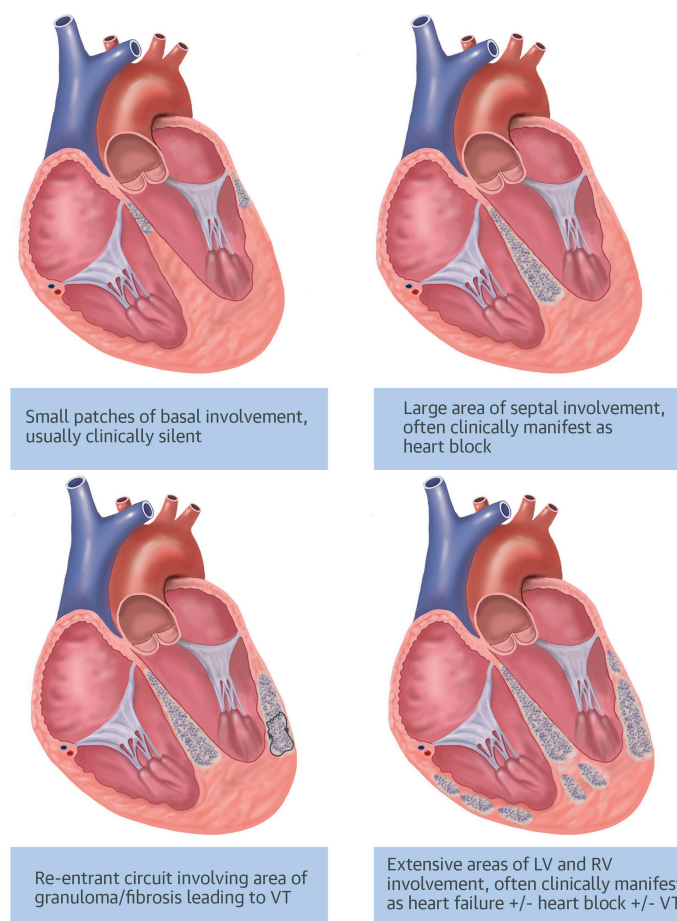


Figure 3 | Clinical features of cardiac sarcoidosis.

Top left panel: Small patches of basal involvement, usually clinically silent disease. Top right panel: Large area of septal involvement that often clinically manifests as heart block. Bottom left panel: Reentrant circuit involving an area of fibrosis or granuloma leading to ventricular tachycardia. Bottom right panel: Extensive areas of left and right ventricular involvement that often clinically manifest as heart failure, heart block, and/or ventricular tachycardia. (Reprinted by permission from Elsevier: *Journal of the American College of Cardiology*, “Cardiac Sarcoidosis” by Birnie DH, Nery PB, Ha AC, et al. Copyright, 2016.)

1.4 SARCOIDOSIS DIAGNOSIS

The diagnosis of sarcoidosis is a challenging and time-consuming process due to the systemic nature and heterogeneity of the disease and the lack of a pathognomonic test or examination. The diagnostic procedure often involves a compatible clinical picture of constitutional symptoms in line with systemic inflammation combined with symptoms and signs originating from the affected organ or organs, a pathological X-ray, histological confirmation of sarcoid granulomas in biopsies, and the exclusion of several other diseases of infectious or non-infectious etiology with similar symptoms and/or histological findings.

The probability of sarcoidosis diagnosis increases dramatically in the presence of clinical findings that suggest one of the syndromic clinical phenotypes of sarcoidosis [25]. As previously mentioned, bilateral hilar lymph node enlargement on chest imaging, erythema nodosum skin lesions, ankle periarticular inflammation combined with recent onset of fever highly suggest the presence of Löfgren's syndrome. Similarly, fever combined with parotitis and uveitis is the classical triad needed for the diagnosis of Heerfordt's syndrome [1]. In most patients, however, symptoms and signs of disease are not highly specific, and a battery of often invasive examinations are required to set the diagnosis. These patients will have to undergo radiographic imaging, usually a chest X-ray and/or a high-resolution computed tomography of the lungs, bronchoalveolar lavage, and biopsy of a suspected area of granulomatous inflammation [1].

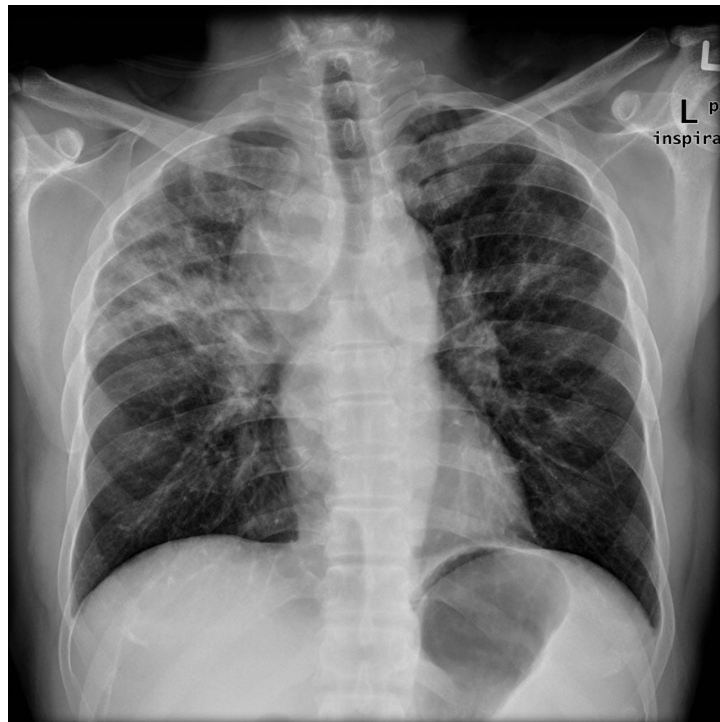


Figure 4 | Stage II sarcoidosis on chest radiograph.

Posteroanterior chest radiograph (X-ray) showing hilar nodal enlargement and lung parenchymal lesions.
Case courtesy of Associate Professor Frank Gaillard, Radiopaedia.org, rID: 6546 (CC BY-NC).

A chest X-ray (Figure 4, page 5) is the examination of choice for suspected pulmonary sarcoidosis in many settings. Common findings on a chest X-ray include the widening of the mediastinum, symmetrical enlargement of hilar lymph nodes and bilateral nodular shadows spreading from perihilar regions to the upper-lung periphery, with signs of fibrosis becoming apparent in advanced disease [26,27]. High-resolution computed tomography is surpassing chest X-rays in popularity due to continuously lowering costs, increased availability and diagnostic utility. Although there are no highly specific findings for sarcoidosis, the presence of symmetrical mediastinal lymphadenopathy, nodular lesions along bronchovascular bundles in subpleural regions in the upper and middle fields, or features of pulmonary fibrosis are suggestive of sarcoid inflammation [26–28]. Other imaging techniques, such as positron emission tomography combined with computed tomography and magnetic resonance imaging are useful in patients with extra-pulmonary disease, e.g., cardiac sarcoidosis [27].

Table 1 | Differential diagnosis of sarcoidosis.

Disease	Findings diverging from sarcoidosis diagnosis	Primary localization of disease
<i>Infectious</i>		
Tuberculosis	Positive culture for mycobacteria	Lungs, lymph nodes
Nontuberculous mycobacterial infection	Positive culture for mycobacteria	Lungs, lymph nodes
Aspergillosis	Positive culture, serology, or histology	Lungs
Histoplasmosis	Positive culture, antigen found in urine	Lungs, lymph nodes
Herpes zoster infection	Granulomas in biopsy	Skin, lungs
<i>Toxoplasma gondii</i> infection	Positive serology	Lymph nodes, skin, liver
<i>Non-infectious</i>		
Lymphoma	Clonal cell expansion	Lymph nodes
Sarcoid-like reaction to tumor	Primary tumor	Lymph nodes
Common variable immunodeficiency	Recurrent infections, hypogammaglobulinemia	Lungs, lymph nodes
Inflammatory bowel disease	Gastrointestinal symptoms and endoscopic findings	Gastrointestinal tract, lungs
IgG4-related disease	Elevated serum IgG4, granulomas rare	Lungs, lymph nodes
Hypersensitivity pneumonitis	History of exposure to organic matter, poorly formed granulomas	Lungs, lymph nodes
Berylliosis	History of exposure to inorganic particles	Lungs, lymph nodes
Drug-induced granulomatous disease	History of exposure to interferon β , tumor necrosis factor α , or checkpoint inhibitors	Lungs, lymph nodes

Adapted from [27].

In patients with pulmonary involvement, for whom biopsy of skin or visceral organs is not indicated, tissue obtained during bronchoscopy (via ultrasound guided transbronchial needle aspiration or other endoscopic procedures) positive for sarcoid granulomas can help set the diagnosis in about 80% of patients with pulmonary sarcoidosis [25]. In addition, analysis of bronchoalveolar lavage fluid showing more than 25% lymphocytes and a CD4 to CD8 T cell ratio higher than 3.5 further increase the diagnostic yield of bronchoscopy [25].

Pulmonary function testing (i.e., spirometry and diffusing capacity for carbon monoxide assessment) alone or in combination with a six-minute walk test are an important component of the initial evaluation of an individual with sarcoidosis [29,30]. In most individuals with sarcoidosis, pulmonary function may be normal at presentation [29]. Abnormal findings in some patients, for example, reduced diffusing capacity of the lung, are associated with severe disease and are therefore decisive for administering immunosuppressive treatment [29]. Because these examinations are standardized and thus reproducible, they are also useful for following-up patients during the course of their disease [29,30].

During all stages of diagnosing a suspected sarcoidosis case, from history taking to analysis of biopsied tissue, alternative diagnoses should be considered. Table 1 (page 6) provides a list of common infectious and non-infectious diseases that mimic sarcoidosis and its findings. Those diseases should be excluded to increase the specificity of sarcoidosis diagnosis.

Several markers of granulomatous inflammation have been tested in the pursuit of identifying one (or more) that could aid the diagnosis of sarcoidosis or determine disease activity during follow-up of patients. Although a few are used today to support diagnosis, none is proven to be adequately sensitive and specific for sarcoidosis. Examples include angiotensin-converting enzyme, serum and urine calcium, soluble interleukin-2 receptor, chitotriosidase, and high molecular weight glycoprotein Krebs von den Lungen-6. Of those, angiotensin-converting enzyme is the most prominently used worldwide [2,31–34]. Angiotensin-converting enzyme is abnormally produced by epithelioid cells in the sarcoid granuloma and is found increased in serum in more than 50% of patients around the time of diagnosis [31,35]. However, several factors markedly limit its usefulness in sarcoidosis. Angiotensin-converting enzyme lacks specificity for sarcoidosis, its levels in serum are affected by common inhibitory medications and do not respond to immunosuppressive treatment used to treat sarcoidosis [31,32,36]. New hope in identifying better biomarkers for sarcoidosis arose from recent discoveries implicating signaling pathways such as the Janus kinase-signal transducer and activator of transcription and mammalian target of rapamycin in granuloma formation [37–39]. It remains to be seen how these novel findings will translate into biomarkers that can be used in daily clinical practice to prognosticate the course of sarcoidosis.

Disease severity indices are useful tools in the clinic to guide treatment and follow-up of patients, and in research studies to predict adverse outcomes. Despite the emergence of modern imaging and molecular techniques and that of statistical methods that can combine multiple layers of data, there is no established index available for sarcoidosis. Some attempts have been made to develop severity scores for sarcoidosis [40–43], but the high heterogeneity in clinical

picture and natural history, differences in disease phenotypes across the globe, and the lack of validation studies limited the clinical and research utility of these scores.

Table 2 | Scadding stages of pulmonary sarcoidosis.
Staging of pulmonary sarcoidosis on chest X-rays (posteroanterior view) and its association with expected spontaneous resolution of disease as suggested by Guy Scadding in 1961 [44].

Stage	Findings on chest radiograph	Frequency at presentation	Expected spontaneous resolution in five years
0	None	5–15%	—
I	Hilar enlargement of lymph nodes	25–65%	60–90%
II	Hilar enlargement of lymph nodes and lung parenchymal disease	20–40%	40–70%
III	Lung parenchymal disease only	10–15%	10–20%
IV	Pulmonary fibrosis (end-stage lung disease)	5%	0%

A staging system for pulmonary diagnosis based on chest X-rays developed by Dr. Scadding in the United Kingdom in 1961 [44] is still used today to prognosticate disease in both the clinic and research studies. Despite its prominence, the Scadding staging system (Table 2) is widely criticized due to the low interobserver variability, weak association with important outcomes for patients (e.g., mortality and quality of life), or pulmonary function, and its inability to adequately inform treatment choice for pulmonary sarcoidosis [45,46].

1.5 SARCOIDOSIS TREATMENT

There is no cure for sarcoidosis. Sarcoid inflammation appears to resolve in about 60% of individuals within two to five years from disease diagnosis with or without pharmacologic treatment [2]. The approach to sarcoidosis treatment is not guided by evidence-based guidelines due to the lack of well-designed experimental or observational studies on large enough patient cohorts. Treatment recommendations are, therefore, based on regional or international consensus [47,48] and treating physicians' preferences. Sarcoidosis treatment is usually divided into two components: symptomatic and immunosuppressive.

As the name implies, symptomatic treatment is administered to alleviate a range of symptoms seen in various clinical phenotypes of sarcoidosis. Non-steroidal anti-inflammatory drugs are considered in individuals with fever and other constitutional symptoms, or periarticular inflammation in patients with Löfgren's disease [4]. Similarly, inhaled corticosteroids may reduce cough and dyspnea, common pulmonary manifestations of sarcoidosis, and pulmonary rehabilitation can alleviate fatigue and dyspnea, increase the working capacity and generally improve the quality of life in patients with sarcoidosis [49,50].

Most sarcoidosis experts advocate immunosuppressive treatment when symptoms and signs of sarcoid inflammation markedly impact a patient's quality of life, as is the case for patients with severe dyspnea and/or fatigue, or when disease progressively impacts vital organs or systems such as the heart, the eyes, or the central nervous system or increases the risk of premature death (a dogma known as the "Wells law") [1,47,51,52]. It should be noted that no widely accepted definitions exist for the concepts of 'great impact' on quality of life or 'danger' due to progressive disease. Three lines of pharmacologic immunosuppressive treatment are available, which function to confine the underlying inflammatory process [1].

Systemic (oral) corticosteroids are the first choice and are often initiated in patients with debilitating symptoms or signs of disease progression [53,54]. A common treatment regime is to prescribe 30–40 mg prednisolone daily for four weeks. and based on the response, the dose is tapered by 10 mg every four weeks [2,55]. In most cases, six to 12 months of treatment with 5–10 mg prednisolone daily will be required to control disease activity and symptoms [55]. Lack of treatment response within three to six months should prompt the use of alternative, second line treatments [2,47]. Adverse events including weight gain, diabetes mellitus, muscle weakness, glaucoma, cataract, and osteoporosis are expected in patients treated with corticosteroids [56,57]. Due to their unfavorable safety profile, tapering and initiation of second line treatments is therefore advised after the initial response to corticosteroids [53,55].

The most popular second line choices in Sweden are methotrexate and azathioprine [54,58]. Methotrexate is administered orally in a dose of 7.5–15 mg per week followed by 5 mg folic acid 24 to 48 hours after methotrexate administration [55]. Azathioprine is given daily in a dose of up to 150 mg [55]. Both medications require a longer time than corticosteroids (three to six months) to induce remission of sarcoid inflammation [59]. In general, methotrexate and azathioprine have comparable effectiveness and well-known toxicity [60], but which of the two is of superior effectiveness and safety for use in sarcoidosis is unknown [61]. Third line treatments consist of tumor necrosis factor α inhibitors and are reserved for refractory cases [62,63]. In end-stage cardiopulmonary disease, lung and heart transplantation remains the intervention of choice to improve a patient's prognosis and quality of life [3].

1.6 CLINICAL COURSE OF SARCOIDOSIS

Emerging evidence from large epidemiologic databases suggests that the inflammatory process is covertly ongoing for months to years before symptoms onset and the diagnosis of sarcoidosis is set. Compared to general population controls, individuals who are eventually diagnosed with sarcoidosis are more likely to have contacts with healthcare and to be dispensed medications (Figure 5, page 10), take leave of absence from work due to illness, and have reduced work ability [64–66]. In addition, ocular sarcoidosis, a difficult to diagnose clinical phenotype of sarcoidosis, appears to be associated with delayed diagnosis and treatment of systemic

disease [67]. It remains unknown whether longer preclinical disease is associated with more severe disease at presentation or an unfavorable clinical course of disease in these patients.

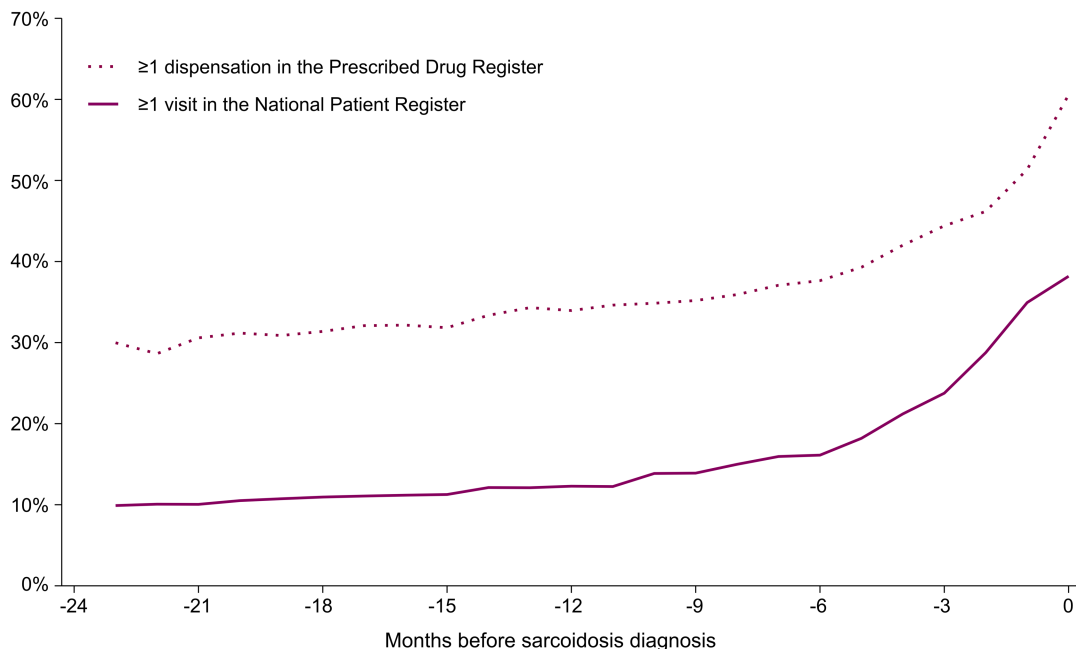


Figure 5 | Healthcare use before sarcoidosis diagnosis in Sweden.

(Adapted by permission from Elsevier: *Respiratory Medicine*, “Sarcoidosis diagnosis and treatment in Sweden: A register-based assessment of variations by region and calendar period” by Rossides M, Kullberg S, Eklund A, et al. Copyright, 2020.)

After diagnosis, sarcoid inflammation and its symptomatology are expected to subside irrespective of treatment in more than half of patients within two to five years [1]. This clinical course of sarcoidosis is often referred to as ‘acute’ and/or ‘self-resolving’ disease. The vast majority of patients with Löfgren’s syndrome belong to this group [16]. In the remaining 30–40% of individuals with sarcoidosis, the disease progresses into a chronic form with remission becoming less likely after five years since diagnosis [68–71]. Organ decline due to fibrosis, sarcoidosis-associated pulmonary hypertension, and the debilitating symptoms (e.g. dyspnea and fatigue) impair patients’ functional and social well-being and increase the risk of premature death [2,68]. As previously mentioned, there is no valid disease severity score and/or biomarker that can reliably predict disease prognosis and the need for treatment or other interventions in patients with sarcoidosis [68,71].

2 LITERATURE REVIEW

2.1 EPIDEMIOLOGY OF SARCOIDOSIS: AN INTRODUCTION

The epidemiology of sarcoidosis, especially the investigation of epidemiologic risk factors and longer-term outcomes in individuals diagnosed with the disease, is the main focus of this thesis. This section begins with a brief introduction of sarcoidosis cohorts used to study sarcoidosis. A discussion of how and why the incidence and prevalence of the disease varies greatly between and within countries and populations follows. The section continues with a depiction of the state of knowledge on select risk factors (genetic or familial and environmental) and outcomes in individuals with sarcoidosis.

Historically, the epidemiology of sarcoidosis was studied using small cohorts of patients seen in specialist clinics or academic centers. The wealth of clinical information collected in these local, often hospital-based cohorts provided some useful insight into patient characteristics, and the diagnosis and treatment of sarcoidosis. However, small numbers and lack of generalizability limit the use of these studies to answer broader research and clinical questions on disease etiology and the course of disease. Large data sources that can be used to study the epidemiology of sarcoidosis are needed.

Few of those are available. Table 3 (page 12) summarizes the characteristics of some of these resources that were used to study risk factors for sarcoidosis and outcomes of patients with the disease in the past two decades. A potential disadvantage of larger databases that rely on self-reports or International Classification of Diseases (ICD) codes to identify sarcoidosis compared to smaller clinical cohorts is that the first are lacking in terms of detailed clinical information that could facilitate validation of diagnoses and allow stratification of patients by clinical phenotype.

2.2 SARCOIDOSIS OCCURRENCE

Sarcoidosis is relatively rare and varies greatly by ethnicity. In Sweden, the incidence and prevalence of sarcoidosis are among the highest worldwide [72,73], surpassed only by those observed in black American populations [74,75]. About 1100 individuals are diagnosed with the disease annually in Sweden and more than 60 000 individuals had a history of the disease in 2013 [76]. Table 4 (page 13) shows estimates of sarcoidosis incidence and prevalence from around the world. It should be noted that recent estimates of incidence and prevalence in most countries of the world are not available due to the lack of large population-based epidemiologic studies. In addition, comparison of available estimates among countries is hindered by the profound variation of methods used to ascertain disease in individual studies (i.e., ICD-coded register data, hospitalization data, insurance claims data, mass radiographic surveys, etc.).

Table 3 | Data sources used in studies of sarcoidosis epidemiology around the globe.

Study name	Country	Study coordinator	Study design	Enrollment period	Sample size	Sarcoidosis ascertainment method	Comparators	References
A Case-Control Etiologic Study of Sarcoidosis (ACCESS)	United States	National Heart, Lung, and Blood Institute	Case-control cohort	1996–1999	720	Adults with biopsy-positive sarcoidosis or Kveim-positive erythema nodosum enrolled from 10 medical centers	Identified through random digit dialing and matched 1:1 on age, sex, and ethnicity. Individuals with granulomatous disorders excluded	[77]
Olmsted County, Minnesota	United States	Rochester Epidemiology Project	Case-control cohort	1976–2013	345	ICD codes, validated through medical chart review; cases were biopsy-positive	Age- and sex- matched comparators sampled from the Project	[78]
Black Women's Health Study	United States	Boston University, Slone Epidemiology Center	Cohort	1995–2011	454	Self-reported	Study participants who did not report sarcoidosis	[79]
Nurses' Health Study II	United States	Brigham and Women's Hospital, Boston, Massachusetts	Cohort	1989–2011	261	Self-reported	Study participants who did not report sarcoidosis	[80]
Taiwan National Health Insurance Research Database	Taiwan	Ministry of Health and Welfare	Case-control	1997–2015	>2000	ICD codes	Age- and sex-matched controls	[81,82]
Optum database	United States	Optum, UnitedHealth Group	Cohort	2009–2013	6831	≥ 2 ICD codes, ≥ 14 days apart	Non-sarcoidosis Optum members	[74]
Health Improvement Network database	United Kingdom	Health Improvement Network	Cohort	1991–2003	1019	Read codes from primary healthcare practices	Non-sarcoidosis comparators in the database matched 4:1 on age, sex, and general practice	[83]
Swedish sarcoidosis linkage	Sweden	Clinical Epidemiology Division, Karolinska Institutet	Case-control cohort	1964–2013	>10 000	ICD-coded visits in the National Patient Register	Non-sarcoidosis comparators matched 10:1 on birth year, sex, and residential location	[84,85]

ICD = International Classification of Diseases.

Most individuals receive their diagnosis of sarcoidosis in middle age; pediatric disease is rare [1,72,86]. Sarcoidosis affects men and women in nearly equal numbers, but the incidence peaks in the mid-thirties for men and a decade later in women [6,72,76]. Differences in age of onset among females and males have been observed in numerous studies worldwide, especially those based on population-based samples [72]. The reasons remain unknown. One could speculate that certain hormonal factors (e.g., sex hormones) may delay the onset of sarcoid inflammation or disease symptoms in females, albeit all but one study in females has examined those [87]. Investigators found that longer exposure to endogenous sex hormones was associated with a lower incidence of sarcoidosis in black women from the United States [87]. No effect on sarcoidosis incidence was observed with oral contraceptive use [87].

Table 4 | Sarcoidosis incidence and prevalence from around the world.

Country (territory)	Annual incidence^a	Prevalence^a	References
Australia (Victoria)	4.4	—	[88]
Canada (Ontario)	6.8	143	[89]
Denmark	6.4	—	[90]
Finland	11.5	28	[91]
France (Greater Paris)	4.9	30	[92]
Greece	1.1	6	[93]
Italy (Parma)	—	49	[94]
Japan ^b	1.0	5	[91,95]
Poland (Silesia)	7.1	6	[96]
South Korea	1.3	3	[97]
Sweden	11.5	160	[76]
Switzerland	7.0	121	[98]
United Kingdom	5.0	—	[83]
United States	8.3	60	[74]

^aEstimates are per 100 000 individuals.

^bIncidence for the whole of Japan; prevalence in Hokkaido.

Within-country geographical variation in disease prevalence is a worldwide phenomenon [80,94,99] and Sweden is no exception. Historically, a higher prevalence of sarcoidosis is found in the northwestern parts of the country irrespective of method of ascertainment, either through radiographic surveys from the 1940–50’s [100] or register data in the 2000’s [76]. The reasons for this clustering of cases remain unclear. A recent study indicated differences in sarcoidosis diagnosis and treatment among different Swedish regions, which may suggest differences in awareness of the disease and/or variation in sarcoidosis severity within the country [58]. However, those alone cannot explain the difference in disease occurrence observed in the country as differences in diagnosis and treatment do not appear to match the differences in prevalence [58]. Last, clustering of cases during winter or late spring has been described in

some cohorts [101,102] and in Löfgren's syndrome [16], but time series analyses did not reveal any seasonal patterns in other data [103].

2.3 RISK FACTORS FOR SARCOIDOSIS

2.3.1 Genetics of sarcoidosis

The etiology of sarcoidosis remains unknown and our understanding of the pathophysiology is limited. An environmental insult is believed to trigger sarcoid granulomatous inflammation in a genetically susceptible individual [1]. The role of genetics, environmental, as well as other modifiable exposures in sarcoidosis is discussed in the lines that follow.

The description of the first clusters of sarcoidosis in families (termed 'familial aggregation') in anecdotal reports gave rise to the notion that genetics are involved in sarcoidosis occurrence [104]. Further support for this hypothesis originated from the failure to identify a transmittable pathogenic agent. Accumulating knowledge on population (quantitative) genetics in the late 1960's allowed us to quantify familial aggregation [105–108]. The first studies were small, descriptive and lacked the rigor of modern well-designed epidemiologic assessments. Nevertheless, by showing that first degree relatives had sarcoidosis in higher proportions than what was expected in the general population, they formalized the idea that genetics is an important parameter in sarcoidosis pathophysiology.

The rise of the new millennium saw a more rigorous quantification of familial aggregation with the estimation of familial relative risks. Familial relative risks, defined as the risk for developing the disease associated with existence of the disease in relatives, are very informative for clinicians who are concerned with the differential diagnosis of sarcoidosis and patients who are concerned with the heredity of their disease. A study from the United Kingdom reported relative risks associated with having a diseased first degree relative (50% genetic similarity) in the range of 36 to 73 [109], higher than the one reported for a black American population (2.5) [110]. The low precision and the differences between these two reports are more likely to have been influenced by small numbers and issues with study design. Information on sarcoidosis was collected for cases' relatives via self-reports at the expense of a higher risk for recall bias. In addition, unreliable estimates of sarcoidosis prevalence were used in the absence of an active control group.

The ACCESS (A Case-Control Etiologic Study of Sarcoidosis) was a landmark American study [111] in which 646 pairs of sarcoidosis cases and population controls were enrolled. Relatives and their sarcoidosis status were self-reported via questionnaires addressed to cases and controls, but disease status was validated for some. ACCESS reported that having a first degree relative with the disease, was associated with a 3.8-fold increased risk for developing the disease [111]. Second degree kinships, in which 25% of the genome is shared, carried a higher relative risk of 5.2 [111]. Considerable effect modification by ethnic background was also observed, with familial relative risks of 3.1 estimated for black American and 16.6 for

white individuals [111]. Despite the larger sample size in the ACCESS study compared to previous investigations, the influence of potential biases was not diminished. These biases include small numbers due to the rarity of sarcoidosis in relatives of controls, an unconventional statistical approach whereby relatives' data was analyzed in a prospective manner ignoring the case-based sampling, and measurement error due to the ascertainment of disease status in relatives via self-report.

Familial aggregation may arise because genes, the environment, or both, are shared among family members. Heritability is a population-level measure that captures the amount of variation in the susceptibility to developing a phenotype that is attributable to genetic factors. It is frequently derived from twin or other family-based studies [112]. Knowledge of the magnitude of the genetic component of a disease may assist the prioritization of research projects and advance diagnosis and treatment. In sarcoidosis, heritability was estimated to be 66% in a small study of Danish and Finish twins (5 concordant twin pairs) [113]. Genetics are thus perceived to be of vast importance in sarcoidosis etiopathogenesis calling for serious efforts to identify those factors by the means of molecular methods. A large genetic contribution to disease etiology also entails that the disease is much less likely to be preventable.

Table 5 | Select genes associated with sarcoidosis.

Candidate gene	Protein	Implication in other diseases
ACE	Angiotensin-converting enzyme	Stroke, diabetic nephropathy
ANXA11	Annexin A11	Amyotrophic lateral sclerosis
BCL2	BCL2 apoptosis regulator	Chronic lymphatic leukemia, non-Hodgkin lymphomas
BTNL2	Butyrophilin-like protein 2	Berylliosis, rheumatoid arthritis, Crohn's disease
IFNG	Interferon γ	Type I diabetes mellitus, aplastic anemia, rheumatoid arthritis, multiple sclerosis
IL18	Interleukin 18	Metabolic syndrome, rheumatoid arthritis, malignant neoplasms
IL23R	Interleukin 23 receptor	Ankylosing spondylitis, psoriatic arthritis, Crohn's disease, ulcerative colitis
NOTCH4	NOTCH receptor 4	Rheumatoid arthritis, systemic lupus erythematosus, schizophrenia
TGFB1	Transforming growth factor β	Idiopathic pulmonary fibrosis, solid tumors
TNFA	Tumor necrosis factor α	Rheumatoid arthritis, systemic lupus erythematosus, psoriasis

Data obtained from the DisGeNET discovery platform (www.disgenet.org).

On the molecular level, no single mutation can explain the susceptibility to sarcoidosis. Alleles of the major histocompatibility complex (classes I and II) as well as other genes have been

associated with sarcoidosis onset [114–120], distinct sarcoid phenotypes [16,116], and the prognosis of sarcoidosis [16]. Particularly, class II alleles HLA-DRB1*1501 and HLA-DRB1*0401 were found to be associated with an increased risk for developing sarcoidosis in Europeans [121]. Scandinavian patients bearing HLA-DRB1*0301 alleles are predisposed to sarcoidosis, particularly to Löfgren’s syndrome, which is more likely to resolve within a few years from diagnosis compared to other disease phenotypes [16].

Select other loci on the human genome that were associated with susceptibility to sarcoidosis in genome-wide association and whole-exome sequencing studies [114–118,120,122–128] are summarized in Table 5 (page 15). Of note, these genes are not unique to sarcoidosis risk; they were implicated in several other complex diseases of mostly autoimmune etiology (Table 5, page 15). Our understanding of the underlying genetics of sarcoidosis is minimal [116]. Larger studies and possibly more sensitive methods are needed. Epigenetics and gene-environment interactions are two parameters that also deserve our attention as they may contribute to our understanding of the complex mechanisms driving sarcoid inflammation [121,129]. Recently, investigators presented findings suggestive of excess risks for sarcoidosis than those expected in genetically predisposed individuals when individuals smoked [130], or separately, if they were exposed to insecticides [131].

2.3.2 Environmental exposures predisposing to sarcoidosis

Similar to genetics, our understanding of the role of environmental, that is, modifiable and likely preventable exposures, is limited. Of those exposures, the most investigated include infectious and occupational agents (metals and other inorganic dusts), and lifestyle factors like smoking and obesity.

2.3.2.1 Infectious agents and diseases

Many believe that sarcoidosis is caused by an infectious agent. This is the oldest hypothesis for sarcoidosis etiology and gained prominence when clinical and histologic similarities between sarcoidosis and tuberculosis (a mycobacterial infection) were realized [132]. Molecular studies that attempted to establish a causal link between pathogen and disease are numerous [133]. As shown in Table 6 (page 17), a wide spectrum of bacteria [134–136], viruses [137,138], and fungi [139,140] were examined over the years. *Mycobacterium tuberculosis* and *Cutibacterium acnes* (formerly *Propionibacterium acnes*) are the most prominent pathogens associated with sarcoidosis occurrence [133]. The precise mechanisms by which these microorganisms trigger and sustain the granulomatous inflammatory process in sarcoidosis have yet to be fully elucidated.

In recent years, analysis of the human microbiome is thought to be a promising new way of identifying patterns that could relate to immune dysregulation, disease onset, and disease progression. In sarcoidosis, studies that analyzed patients’ lung microbiota could not identify any disease-specific patterns that could distinguish patients with sarcoidosis from healthy

controls and/or individuals with other interstitial lung diseases [141–143]. One such study indicated *Atopobium* spp. and *Fusobacterium* spp. as promising candidates that merit further investigation, but could not identify any notable differences in the abundance of *Mycobacterium* spp. or *Cutibacterium* spp. among patients with sarcoidosis, idiopathic pulmonary fibrosis, or healthy controls [143].

A cause and effect relationship between a pathogen and sarcoidosis onset is difficult to infer with confidence from molecular studies due to the lack of replication of most findings [132,144]. Criticisms against most molecular studies pinpoint to the lack of standardization of extraction and identification techniques, and importantly, to the fact that remnants of pathogens were isolated from individuals already suffering from the disease.

Table 6 | Select infectious agents implicated in sarcoidosis occurrence by molecular studies.

Microorganism	Perceived evidence	References
<i>Mycobacterium tuberculosis</i>	DNA and other molecular remnants, particularly the mycobacterial catalase-peroxidase protein identified in sarcoidosis lesions, but <i>Mycobacterium tuberculosis</i> was not cultured. Disease activity dependent T cell response in blood and bronchoalveolar lavage fluid was demonstrated	[145–147]
<i>Cutibacterium acnes</i>	DNA and other molecular remnants isolated from sarcoid lesions (especially lymph nodes) and cultured. Immune responses to <i>C. acnes</i> were demonstrated. Most studies involved Japanese patients; studies in other patient populations are limited	[135,148,149]
<i>Borrelia</i> spp.	DNA and/or protein remnants isolated	[150,151]
<i>Human herpesvirus 8</i>	Viral DNA isolated in one study	[152]
Fungi	Fungal exposure associated with immunologic responses in sarcoidosis. Antifungal antibodies isolated from serum and bronchoalveolar lavage fluid of patients	[139,140,153]

Direct and indirect epidemiologic evidence of the role of infectious pathogens in sarcoidosis onset is limited. From studies on disease distribution, especially those which demonstrated significant geographical and/or seasonal variation in sarcoidosis occurrence, we can infer that infectious agents may contribute to disease onset. In addition, evidence from small studies is suggestive of the role of antimicrobial medications as potential treatments of sarcoidosis. The concomitant levofloxacin, ethambutol, azithromycin and rifampicin regimen (“CLEAR”) showed promising results for cutaneous sarcoidosis in a small study [154]. It is currently being tested in the United States in a larger placebo-controlled clinical trial focused on pulmonary disease [155]. Similarly, the efficacy of antimicrobial medications (in addition to standard corticosteroid treatment) for cardiac sarcoidosis is being evaluated in a large trial in Japan [156]. It should be emphasized, however, that it remains unclear whether any potential benefits of antimicrobial treatment in sarcoidosis are a result of pathogen elimination or simply due to immunomodulatory effects exerted by medications of this class [157,158].

In previous years, the scarcity of prospectively collected data available for research largely limited our potential to answer these fundamental etiologic questions. Only recently, a register-based study from Taiwan indicated that a diagnosis of tuberculosis was associated with an eightfold increased rate of receiving a sarcoidosis diagnosis in the future [82], although estimation was based on a small number of cases.

2.3.2.2 *Occupation and occupational exposures*

Several occupations and occupation-attributed exposures were associated with an increased risk for sarcoidosis in case-control studies (Table 7). Among the implicated exposures are inhalable hazardous agents such as insecticides, metals, and inorganic dusts [159]. These toxic agents are found in abundance in the work environment of miners [160], firefighters [161], and agricultural workers [162]; all of which are occupational groups with increased sarcoidosis incidence compared to the general population.

Table 7 | Occupations associated with increased risk of sarcoidosis.

Occupation	Suspected agents	References
Firefighters	Inorganic dust, particulate matter	[163]
Emergency response personnel	Inorganic dust, particulate matter	[163]
Military personnel	Inorganic dust, mold	[164]
Office workers	Toner, mold	[162]
Miners	Metal, inorganic dust, silicates	[162]
Metallurgy/foundry workers	Metal, silica, inorganic dust, fluid aerosols	[160,165]
Construction workers	Inorganic dust	[162]
Farmers	Insecticides, vegetable dust, non-public water	[162,166]

2.3.2.3 *Lifestyle factors: smoking and obesity*

A protective effect for smoking has been suggested in several case-control studies [162,167,168]. Ever smoking was associated with a 35% decreased odds for sarcoidosis [162] and current smoking reduced the risk by 66% [168]. These findings may have been influenced by selective recall of smoking habits in sarcoidosis cases compared to controls. However, a protective effect for ever smoking (hazard ratio 0.5) was also shown in a Swedish cohort study of construction workers [169], which is much less likely to have suffered from recall bias. Whether this controversial association is true or merely an epidemiologic artifact is difficult to disentangle. It is worth noting that smoking was not found to decelerate disease progression [170]. Nonetheless, smoking was found to induce downregulation of adaptive immune responses and thus protect against disease [171].

High body mass index has been associated with an increased risk for sarcoidosis in several studies from the United States and Scandinavia, most of which focused exclusively on females. In a series of prospective studies utilizing the Black Women's Health Study, a 42% to 74%

higher risk for developing sarcoidosis was observed both when obesity (body mass index ≥ 30 kg/m²) was ascertained at age 18 and two years prior to sarcoidosis diagnosis [79,172]. Similarly, males and females with sarcoidosis from Olmsted County, Minnesota, were 2.5-times more likely to be obese around diagnosis than controls [168]. Obese pregnant women in Denmark were four times more likely to be diagnosed with sarcoidosis in the future than women with normal pre-pregnancy body mass index [173]. Higher risks for sarcoidosis associated with obesity persisted in all studies despite adjustment for other lifestyle and socioeconomic determinants. Dysfunction of adipose tissue is thought to influence immune regulation in the lung [174], but the processes linking obesogenic inflammatory pathways and sarcoid granuloma formation are overall poorly understood.

2.4 OUTCOMES OF PATIENTS WITH SARCOIDOSIS

In the eyes of some, physicians included, sarcoidosis is considered a benign disease; a long lasting flu [175]. It may be true that a large group of patients, especially those with Löfgren's syndrome, will exhibit a self-limiting disease lasting for two to five years [1]. For a smaller but significant group, however, sarcoid inflammation does not self-resolve despite treatment, resulting in fibrosis and the decline of vital organ function [1]. The existence of various disease phenotypes and the lack of severity scores are two of the most important challenges in the study of long-term patient outcomes in sarcoidosis. Despite these challenges, identifying whether sarcoidosis patients overall, or patient subgroups specifically, are at risk of long-term unfavorable outcomes is a prerequisite to apply prevention efforts or alter recommendations for care in these patients.

2.4.1 Mortality

The most critical outcome for sarcoidosis patients is death. Judged by design, studies that have examined the risk of death in individuals with sarcoidosis fall into two very different categories: cross-sectional assessment of death certificate data and prospective studies of mortality. Several European and American cross-sectional assessments of mortality utilizing death certificates have been published over the course of several years. Recently, a mortality rate of 3.6 per million was estimated using French data sources [176], whereas slightly higher rates were reported in somewhat older studies that utilized British [177] and American [178] databases (4.2 and around 5.0 per million, respectively). Most assessments concluded that sarcoidosis mortality increased significantly in the latest years [176–178], but were inconsistent as to which gender experienced the greatest burden.

Results from death certificate studies are highly dependent on the trends of reporting causes of death. This may be problematic because causes of death compete for a place on the death certificate, reporting of certain causes may be incentivized in some countries (including Sweden), and classification systems evolve limiting comparability across studies or calendar

periods. In addition, because sarcoidosis might resolve years before death and thus never reported, it is unclear whether this method of assessing mortality can fully capture the true burden of disease [179]. For those reasons, prospective cohort studies feature a superior design for examining mortality.

Three prospective studies in sarcoidosis were conducted to ascertain the relative risk of death associated with sarcoidosis. They used data from a British electronic primary care database [83], the Black Women's Health Study [180] and a smaller cohort from Olmsted County, Minnesota [78] in the United States. In the first two, sarcoidosis was associated with a twofold increased risk for all-cause mortality [83,180], whereas in the latter smaller study a standardized mortality ratio of 0.9 was reported [78]. In contrast to all death certificate studies, no prospective study observed an increase in the number of sarcoidosis-related deaths in recent years [78,83].

Except for older age, there is no agreement as to which variables can accurately predict premature death in individuals with sarcoidosis [78,83,180]. Identifying the reasons behind the extensive discrepancies among studies is challenging. Low power, variability in the definition of sarcoidosis (self-reported versus biopsy-confirmed), and unmeasured confounding are likely contributors. Problems with previous studies highlighted the need for larger and better designed studies that will consider the variability among various sarcoidosis phenotypes.

2.4.2 Serious infections

Infectious diseases severe enough to lead to hospitalization are termed 'serious'. Serious infections, especially if recurrent, may lead to high healthcare costs and negatively affect a patient's quality of life [179,181,182]. Several factors including, but not limited to, the immune dysregulation associated with sarcoidosis and the immunosuppressive treatment administered in some patients may lead to an increased risk for infection in individuals diagnosed with the disease. Indeed, more than 5% of sarcoidosis patients are hospitalized at least once for infection during the course of their disease [183], and in some cases, chronic aspergillosis may superimpose advanced pulmonary sarcoidosis leading to increased mortality [184]. Nevertheless, information about serious infection risks in sarcoidosis and how those vary by phenotype and during the course of disease is limited.

A twofold increased risk for serious infection was observed in individuals with sarcoidosis compared to the general population in the small cohort from Olmsted County, Minnesota [185]. The risk was 99% higher for pneumonia, which was the most common serious infection in these patients [185]. In the same study, individuals with sarcoidosis who had ever used a range of immunosuppressive treatments during the 18-year-long follow-up of this study were at an even higher risk for developing a serious infection [185]. In a separate study from the United States, patients hospitalized for herpes zoster were 52% more likely to report sarcoidosis as a comorbidity than patients hospitalized for other reasons [186], highlighting a possible association between sarcoidosis and herpes zoster occurrence.

2.4.3 Cardiovascular disease and heart failure

Cardiovascular disease is an umbrella term used to cover diseases affecting the heart vasculature. It is often used more generally to refer to any disease of the arterial vasculature of the whole body (including the central nervous system and the periphery), or their consequences. Atherosclerosis is the most common pathophysiologic mechanism underlying these diseases. Herein, 'cardiovascular disease' is used in its narrow sense to refer to ischemic heart disease and/or acute myocardial infarction. Heart failure, as the term explicitly implies, refers to failure of the heart to fulfill its role, that is, pumping blood in the small and large circulations to maintain adequate flow to cover the needs of body tissues. In the general population, cardiovascular disease and heart failure are leading causes of morbidity and mortality [187]. Accelerated atherosclerosis as a result of the inflammatory processes that characterize sarcoidosis may further exacerbate risks for cardiovascular disease [188]. Little is known, however, whether these pathways translate to observable clinical risks for acute myocardial infarction, and later, heart failure in sarcoidosis.

Three studies examined the risk for cardiovascular disease in sarcoidosis compared to general population but, likely due to small numbers and differences in outcome definitions, results were conflicting. An older register-based study from Sweden indicated a 15% higher risk of acute myocardial infarction in individuals hospitalized for sarcoidosis compared to the general population [189]. It should be noted that only about 9% of the total sarcoidosis cases are diagnosed in inpatient care in Sweden [58], a fact that greatly limits the generalizability of these findings. More recently, a study conducted using data obtained from the Olmsted County, Minnesota cohort from the United States showed a 65% increased risk of cardiovascular disease (defined as a composite outcome) in sarcoidosis compared to the general population, which was mostly driven by increased rates for congestive heart failure [190]. Although higher risks appeared to be associated with corticosteroid use in sarcoidosis [190], limited power did not allow for clear interpretations. Another investigation utilizing primary care data from the United Kingdom, suggested that the rate of myocardial infarction was 40% lower than that in the general population [191]. Stratifying by sex, however, sarcoidosis was found to be associated with a higher relative risk of acute myocardial infarction in male but not female patients (hazard ratio 1.55 versus 0.89, respectively) [191]. Treatment and other contributors to these high risks, if any, were not considered.

Heart ischemia is a common cause of heart failure in the general population [187]. One would think that if risks for acute myocardial infarction are increased in sarcoidosis then risks for (ischemic) heart failure will follow a similar pattern. To complicate matters further, the myocardial tissue may be infiltrated by granulomatous inflammation in sarcoidosis which may result in heart failure of non-ischemic etiology in surviving patients [17]. Recently, a study using register data from Denmark showed that sarcoidosis is associated with a sevenfold higher risk of heart failure of any etiology during the first year after sarcoidosis diagnosis [192]. Despite growing interest in recent years in sarcoidosis-associated heart failure, no study has investigated its etiology in sarcoidosis or attempted to identify patient groups that are at the highest risk of developing heart failure and potential predictors of these risks.

2.4.4 Other somatic and mental health outcomes

Several studies have examined the role of sarcoid inflammation in the occurrence of cancer. A meta-analysis estimated that sarcoidosis is associated with a slight increase in the risk for invasive cancers (relative risk 1.2) [193]. Relative risks were highest (around 2.0) for skin and hematopoietic malignancies [193]. Cancers were more likely to be identified within four years from sarcoidosis diagnosis [193]. Whether sarcoidosis causes cancer or merely facilitates early diagnosis of some cancers remains to be determined. Recent findings support the latter. In a small study from the United States, history of malignancy was similar in cases and comparators at diagnosis (4.3%) [194], and in a large Danish assessment, relative risks for cancer peaked within three months from diagnosis and then rapidly declined [195]. Nevertheless, no studies have directly examined the influence of the extensive screening that individuals with sarcoidosis undergo around the time of diagnosis.

Mental health outcomes are often neglected in the study of somatic disease. Sarcoidosis is no exception. Only a few descriptive studies focused on mental health deterioration in the form of depression and anxiety. Their findings highlight the large impact that sarcoidosis and the accompanying fatigue, dyspnea, and sleep disturbances have on the mental health of patients and their quality of life [196]. Individuals with sarcoidosis not only score highly on depression and anxiety scales [197], but up to 25–30% fill criteria for the diagnosis of major depression and generalized anxiety disorder [198,199].

2.5 KNOWLEDGE GAPS IN SARCOIDOSIS EPIDEMIOLOGY

In the late 1990's, ACCESS was the largest epidemiologic study on sarcoidosis. Despite the immense efforts to fund and conduct this multi-center study, the failure to fulfil its primary aim to identify the cause of sarcoidosis resulted in disappointment. During the last 20 years, epidemiological methods and data sources developed and expanded. Efforts to identify risk factors for sarcoidosis and outcomes of patients with the disease did not follow a similar pattern; they were scarce, small and rarely generalizable to a larger target population. An up-to-date understanding is needed to facilitate evidence-based clinical practice and further research, either clinical, epidemiologic, or molecular. Data on the familiarity of sarcoidosis or the relative risk of premature death, infectious disease, or other debilitating outcomes that could inform treating physicians and patients and facilitate decision making in terms of diagnostics and treatment remained unknown. In later years, the availability of Swedish health and administrative databases for health research, the ability to link records across various sources and complement those with clinical data, and the development of methods and computational power to analyze data obtained from these large sources provided an unprecedented opportunity to revisit clinical and epidemiological research questions and ask new ones.

3 RESEARCH AIMS

The overall objective of this doctoral project was to investigate important risk factors for and outcomes of patients with sarcoidosis using large population-based studies and data from Swedish registers. Specifically, individual studies in this thesis aimed to answer the following research questions:

1. Is having first and second degree relatives with a history of sarcoidosis associated with a higher risk of developing the disease? If yes, how much of the susceptibility to sarcoidosis in the Swedish population can be attributable to genetic variation in the population?
2. Are infectious diseases associated with a higher risk of developing sarcoidosis in the future?
3. Do individuals with sarcoidosis have a higher risk of all-cause death compared to the general population? Does the relative risk of mortality associated with sarcoidosis vary by age, sex, or sarcoidosis treatment status around the time of diagnosis?
4. Is sarcoidosis associated with a higher risk for first and/or recurrent serious infections? Does the risk differ across groups defined by age at sarcoidosis diagnosis, sex, and treatment status around diagnosis?
5. Is initiating methotrexate as second line treatment for sarcoidosis associated with a higher or lower risk of infection at six months compared to initiating azathioprine?
6. Is the risk of heart failure higher in sarcoidosis compared to the general population? Does the relative risk of heart failure vary by age, sex, sarcoidosis treatment status around diagnosis, time since sarcoidosis diagnosis, or history of ischemic heart disease? Which clinical risk factors best predict a heart failure diagnosis after sarcoidosis occurrence?

4 METHODS

4.1 OVERVIEW OF EPIDEMIOLOGICAL METHODS

This section attempts to provide an overview of the epidemiologic methodology used in this thesis. More focus was directed towards some of the methods and study designs which are either novel or not (yet) commonly used in the literature. More information can be found in the key resources cited in this section.

Epidemiology concerns populations, not distinct individuals. In other words, epidemiology is all about groups of individuals who share a common feature, a disease. Epidemiologic research is quantitative and can be broadly divided into descriptive and etiologic. Descriptive epidemiology aims to quantify the presence of a disease status and/or factors that influence that (e.g., genetic, environmental, or social) in a well-defined population. Etiologic research aims to identify which factors can influence a disease state (either positively or negatively) and how disease characteristics are associated with favorable or adverse outcomes. The ultimate purpose of etiologic research is to understand the biologic and social mechanisms and processes that lead to disease development or govern the course of disease and identify ways to impact those in order to improve health. All six studies in this thesis are etiologic, aiming to answer the question “what is the effect of an exposure on the outcome?”.

Two popular study designs are used in etiologic epidemiologic research: case-control and cohort designs. Their scope is to quantify the impact of an exposure on an outcome. Since allocation of the exposure is not randomized and its administration is not controlled in neither of the two designs, they are collectively thought to be observational designs.

Case-control and cohort studies. In the first two out of six projects in this thesis, I used the case-control study design. These case-control studies were embedded (or ‘nested’) in a defined dynamic population, the whole Swedish population. The term ‘dynamic’ refers to the fact that eligible members of the population change with a unit of time (herein, days), but overall, the population is approximately static over a period as new individuals gain membership while others exit the population pool [200]. From that dynamic population, cases of a disease that arise each day are identified and simultaneously, a number of at-risk individuals without the disease at the time (controls) are randomly sampled and matched to each case. In the literature, this process is referred to as “incidence density sampling” or “matching on calendar time” [201]. Matching further on other important determinants of the risk of the outcome (e.g., age or sex), render the groups more comparable from the outset, thus maximizing the efficiency of analyses [201]. Historical exposures among cases and controls can be then ascertained. The risk of the outcome related to those exposures is compared within matched case-control groups.

In a cohort study, generally two groups of individuals at risk of the outcome of interest are identified: one with a particular exposure and a group without the exposure which is as similar as possible to the first concerning all other characteristics [201]. Exposure is quantified and the two groups are kept under observation for a predefined amount of time until they develop the

outcome of interest [201]. Occurrence of the outcome in the two groups is then compared and inference on the impact of exposure on the outcome is made [201].

Target trial emulation. Experimental randomized controlled clinical trials are the mainstay of comparative effectiveness and safety research. When trials are deemed either unethical, untimely, or unfunded, already collected real-world data can be used to quantify the benefits and harms of interventions [202]. In the field of clinical epidemiologic research, these interventions are often medication regimens. Target trial emulation methods offer a new approach to comparative effectiveness and safety research (i.e., causal inference) that combines the merits of experimental and observational research [202,203]. The method is self-explanatory; a randomized controlled trial designed to answer a well-defined causal inference research question (termed the ‘target trial’) is emulated with observational (unrandomized and uncontrolled) data. As is the case of randomized controlled trials, a study protocol governs all steps of the emulation process [204]. Recent improvements in our understanding of causal inference and the progress of statistical methods allow for causal effects, which would have otherwise obtained from experimental studies, to be inferred using observational studies.

Quantitative population genetics. In contrast to molecular genetics that focus on genes or markers thereof and how these vary among diseased and non-diseased populations, quantitative genetics allow us to make inference on underlying genetics of a disease by estimating variation of the risk of a disease among groups of related individuals. Is sarcoidosis a quantitative, continuous trait? Yes, at least according to the liability-threshold model, a pivotal model in quantitative genetics that allows us to study genetic and environmental contributions to multifactorial diseases like sarcoidosis. One can think of the liability (risk) for developing sarcoidosis in the general population as a continuous, normally distributed trait whereby some individuals with certain genetic makeup pass a hypothetical threshold for developing the disease under the influence of environmental factors [205].

Individuals whose first degree relatives have sarcoidosis are more likely to exceed the threshold for developing sarcoidosis under the assumption of shared genetics (and/or shared environmental factors) that contribute to increased disease risk. The component of the variance attributable to additive (average, non-interactive) genetic effects on the liability scale is called heritability [112]. It is a measure useful for distinguishing among genetic and environmental contributors to disease risk in a population with obvious benefits for guiding public health research and for realizing the value of genetics (inheritance) in prediction of disease risk [205,206].

Target validity and the need for bias analyses. The accuracy of findings produced in an epidemiological study is of utmost importance in epidemiology. That is especially true for causal inference studies (that is, studies of comparative effectiveness or safety of various medical or public health interventions) in which an inaccurate estimate will have profound implications in practice. Whichever task an epidemiological study aims to accomplish, whether related to description, prediction, or counterfactual prediction, the causal inference framework

provides some useful insights on validity and its threats [207,208]. Table 8 (page 28) attempts to summarize issues related to internal and external validity commonly called “systematic errors” or “biases” in epidemiologic jargon (i.e., confounding, selection, and measurement bias). If applicable, Table 8 also provides some examples on how these were tackled in individual studies included in this thesis.

Despite our best efforts to minimize biases by targeting those during study design or in analyses, it is almost impossible to claim that those have been eliminated completely. One way of dealing with bias is to pretend that it does not exist, an admittedly unscientific approach. Another is to acknowledge its presence and qualitatively discuss its perceived mechanism and impact. A third is to employ methods to quantitatively estimate the uncertainty around the target estimate of association owing to the presence of one or more systematic errors.

I opted to use the latter approach in several of the individual studies in this thesis where either design or analysis tweaks and/or data were not available to enable traditional sensitivity analyses. I used probabilistic bias analyses [209,210] in which bias parameters were drawn from probability distributions (reflecting the uncertainty around the bias parameter) to produce a frequency distribution of estimates of association that were ‘corrected’ for differential and non-differential exposure or outcome misclassification and unmeasured confounding, or the combination of the two errors. An added advantage of probabilistic bias analyses is the incorporation of uncertainty due to random error in the final estimate [210]. However, two disadvantages of (probabilistic) bias analyses should be noted. First, results depend on the validity of the values assigned to bias parameters, and second, these simulations are often computationally intensive even when modern computers with multi-core processing power are available.

Table 8 | Target validity in epidemiologic studies: issues, mitigation techniques, and examples from this thesis.

Types of target validity	Assumptions for identifiability of causal effects	Description, potential violations and considerations	Mitigation techniques	Examples from this thesis
Internal validity	Consistency	Precise definition of the treatment/exposure and observation of that in the data	Careful definition of treatment and choice of appropriate data source	Definition of methotrexate or azathioprine initiation in Study V
	Exchangeability	Treatment groups are exactly similar on all measured and unmeasured factors except for treatment status. Violations occur due to confounding bias (existence of common causes of the treatment and outcome) and/or selection bias (conditioning on common effects of the treatment and outcome or their effects)	Use of G [generalized]-methods, inverse probability weighting (for confounding and/or selection bias), stratification (restriction, matching, or regression), difference-in-difference, instrumental variable, or front-door criterion methods	Violations of exchangeability were pertinent in all observational studies in this thesis. Methods used at the design and analysis stages to mitigate those included matching, regression, inverse probability weighting, and G-like methods (targeted maximum likelihood estimation)
	Positivity	Treatment varies within strata of confounders. Violations occur due to data sparsity (random non-positivity) or non-eligibility for treatment assignment (structural non-positivity)	Careful selection of study eligibility criteria	Definition of eligibility criteria in Study V
	Measurement precision	Measurement bias (or misclassification for categorical variables) is the inaccuracy in the estimated effect due to faulty values assigned to exposure, outcome, or confounding variables. Differential measurement refers to the process in which the outcome affects measurement of the exposure or vice versa	Increase in measurement precision	Differential and non-differential misclassification of sarcoidosis in various studies in this thesis
External validity	Generalizability	Applicability of the estimated effect from the study sample to the target population from which the sample originates	Use of a population-based design or a representative data source. Study of effect measure modification, if possible	A population-based design was used and effect measure modification was considered in most individual studies
	Transportability	Applicability of the estimated effect in one population to another. Effect modification, interference, and multiple treatment versions should be considered		Qualitatively discussed in all individual studies

4.2 OVERVIEW OF STATISTICAL METHODS

Epidemiology is a quantitative science and statistical methods are absolutely necessary for the analysis of epidemiologic data. Because we are studying samples obtained from larger “target” populations, utilizing statistical methods allows us to take sampling bias into account when we estimate effects or associations. In this thesis, a range of frequentist statistical analysis methods were used. In principle, the choice of statistical method employed in each individual study was governed by study design (i.e., case-control or cohort study) and type of outcome (e.g., time-to-event, binary, etc.).

Conditional logistic regression. It is commonly used to analyze data originating from a matched outcome-dependent sampling scheme such as in a matched case-control study [211]. Using a logit link function (for the Bernoulli probability distribution), a binary outcome is modelled by one or more explanatory variables conditioning on the number of cases within each stratum defined by the matching set. As it turns out, this conditional log-likelihood function is the same as the partial likelihood function of a stratified Cox proportional hazards regression model (described below) in which a case-control group is assigned its own stratum [212]. This allows for relatively simple parametrization using widely available statistical software [212]. Although Cox regression is used, the association is quantified on the odds ratio scale. However, the estimate can be interpreted as a rate ratio, risk ratio, or odds ratio depending on the sampling procedure that generated the case-control data [200]. In data obtained using incidence density sampling from a dynamic population, the odds ratio corresponds to an estimate of the rate ratio obtained from a cohort study conducted in that population [200].

Cox proportional hazards regression. Cox regression is the most popular choice to analyze time-to-event (survival) data. Based on the hazard function, which refers to event rate at time t conditional on not developing the event of interest by an earlier time ($t-1$), Cox regression allows us to estimate how the baseline (in the absence of exposure) hazard rate of an outcome increases or decreases by a factor (the hazard ratio) in the presence of an explanatory covariate such as the exposure without modeling the baseline hazard [213]. To enable this, Cox regression assumes that hazards are proportional throughout the modelled time scale, an assumption which may not hold in all situations.

Flexible parametric survival modeling. In situations when hazards among strata of a covariate cannot be assumed proportional, or when those are of interest to model, flexible parametric survival models can be used. Royston and Parmar models allow us to model hazards flexibly on the log-cumulative hazard scale using restricted cubic splines [214]. It should be noted that the log-cumulative hazard scale has no epidemiologic interpretation. It is, however, computationally attractive, it allows for common epidemiologic effects on absolute and relative scales (e.g. hazard ratio or survival) to be readily estimated, and for time-dependent effects to be easily modelled [214]. The flexibility of parametric survival models comes at the price of a rigorous process of model specification.

Frailty survival models. The analysis of times to multiple or consecutive events per person should consider that these times may not be independent of each other. Extending Cox proportional hazards and flexible parametric survival models by adding a frailty (random effect) term allows us to model the unobserved within-person correlation among event times [215].

Targeted maximum likelihood estimation. It is a method based on maximum likelihood estimation aiming to estimate effects in causal inference studies. Compared to other causal inference approaches such as inverse probability weighted measures and G-computation techniques, targeted maximum likelihood estimation has favorable statistical and epidemiologic properties that makes it ideal for causal inference derived from observational data [216]. It is doubly robust allowing for consistent estimation even in the presence of bias in either of the exposure or outcome model, especially when combined with data adaptive methods [217,218]. Targeted maximum likelihood are substitution estimators and are therefore more reliable in environments where data is sparse and outliers exist [217]. Last, targeted maximum likelihood estimation employs an additional targeting step in the estimation procedure that, by gaining information from the data, attempts to minimize bias in the estimated parameter at the least expense of variance [216,217].

4.3 HEALTHCARE IN SWEDEN

All studies in this thesis were performed in Sweden using nationwide register data complemented with clinical information collected on a subset of patients with sarcoidosis diagnosed by pulmonary medicine specialists at Karolinska University Hospital in Stockholm who opted to be included in our local clinical research cohort.

The Swedish healthcare system is accessible to all individuals legally residing in the country and is largely funded by the state through taxation. Healthcare is, however, administered locally at the regional level (regions were previously called counties). Individual regions form a larger *healthcare region*, a platform aiming to facilitate co-operation on tertiary and specialized care. During the study period and until today, there are six healthcare regions that are organized around one or more university hospitals (Figure 6, page 31): Stockholm and Gotland (Karolinska University Hospital), Uppsala-Örebro (Uppsala University Hospital and Örebro University Hospital), West (Sahlgrenska University Hospital), South (Skåne University Hospital), Southeast (Linköping University Hospital), and North (University Hospital of Umeå).

In Sweden, individuals with sarcoidosis are commonly diagnosed and treated at public specialized clinics, mostly in an outpatient setting [58]. Hospitalizations for sarcoidosis make up less than 10% of care received by these patients. For reimbursement and research purposes, a person's interaction with the healthcare system (i.e., outpatient visits, hospitalizations, day surgery, etc.) is recorded in large national databases which are administered by the National Board of Health and Welfare (*Socialstyrelsen*). An individual's unique identification number

(*personnummer*) enables linkage of their records across these and other administrative databases creating a wealthy resource for clinical epidemiologic research.

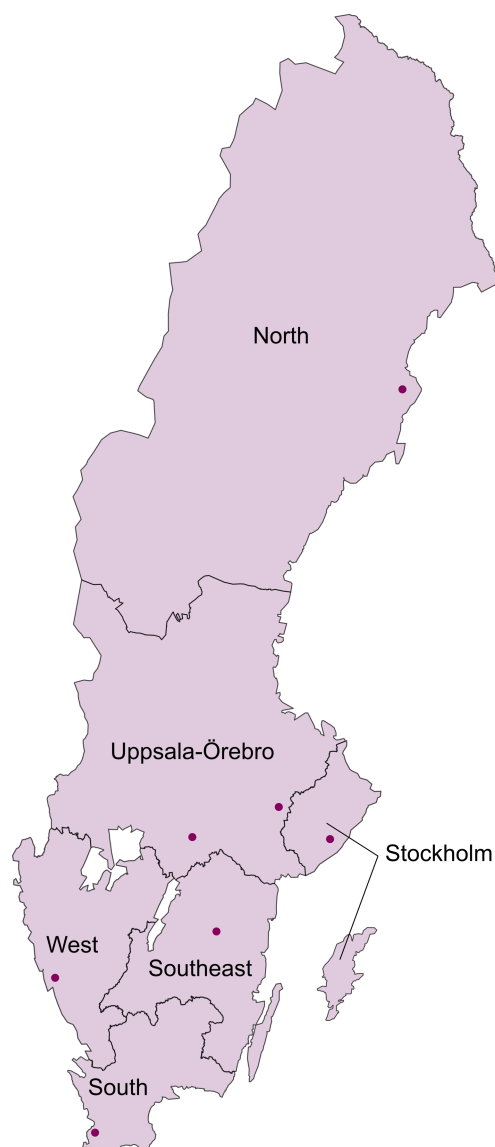


Figure 6 | Swedish healthcare regions.

A dot (●) represents a university hospital (Stockholm and Gotland [Karolinska University Hospital], Uppsala-Örebro [Uppsala University Hospital and Örebro University Hospital], West [Sahlgrenska University Hospital], South [Skåne University Hospital], Southeast [Linköping University Hospital], and North [University Hospital of Umeå]).

4.4 DATA SOURCES

All six studies in this thesis were conducted using data from a linkage of records retrieved from several national healthcare and administrative (sociodemographic) registers held at the National Board of Health and Welfare (*Socialstyrelsen*) and Statistics Sweden (*Statistiska Centralbyrån*).

The following registers were used in the studies in this thesis:

- **National Patient Register** (NPR; *Patientregistret*). First compiled in 1964 to record admissions for in-hospital care, it reached nationwide coverage in 1987 and was further expanded in 2001 to record all visits to outpatient specialist (non-primary care) clinics. Visits have been coded using the Swedish version of the ICD coding system as it evolved and revised throughout the years (ICD-7 1964–1968, ICD-8 1969–1986, ICD-9 1987–1996, ICD-10 1997–today). Each healthcare visit is marked by a date of admission, a date of discharge for hospitalizations, and the main and up to 21 secondary (auxiliary) discharge diagnoses. Day surgeries and other operations are recorded in the NPR since 1997 using the Swedish classification of medical procedures (*Klassifikation av vårdåtgärder, KVÅ*).
- **Prescribed Drug Register** (PDR; *Läkemedelsregistret*). Since July 2005, the PDR holds information on all dispensations of prescribed medications in pharmacies across Sweden. Medication dispensations are coded using the Anatomical Therapeutic Chemical (ATC) classification system and the dates of prescription and dispensation as well as the dispensed amount are available in the register. The PDR does not capture over-the-counter purchases or medications administered in healthcare settings.
- **Cause of Death Register** (*Dödsorsaksregistret*). Founded in 1961, it records dates and ICD-coded causes of death (primary and contributory) for individuals who died either in Sweden or abroad since 1952.
- **Cancer Register** (*Cancerregistret*). Primary malignant tumors that are clinically or histologically diagnosed in hospitals across Sweden are reported to the Cancer Register since 1958. Cancer diagnoses are coded using revisions of the ICD coding system.
- **Total Population Register** (*Registret över totalbefolkningen*). It holds longitudinal demographic and vital data on all individuals residing in Sweden since 1968 including the date of birth, sex, dates of immigration and emigration to/from Sweden, country of birth, residential location, and civil status.
- **Longitudinal Integrated Database for Health Insurance and Labor Market Studies** (*Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier*). By integrating information collected in various administrative registers held at Statistics Sweden and the Social Insurance Agency (*Försäkringskassan*), this database is compiled yearly since 1991 with, among others, data on attained education and income.

- **Multi-Generation Register** (*Flergenerationsregistret*). The Multi-Generation Register keeps data on biologic and, if applicable, adoptive parents of all individuals alive and legally residing in Sweden since 1961 who were born starting 1932 and onwards. Data from the Multi-Generation Register allows for pedigrees to be constructed.

In some of the studies in this thesis, clinical information from the **Karolinska Clinical Cohort** was used to complement the analyses. The cohort includes about 1500 individuals diagnosed with sarcoidosis since the late 1990's by pulmonologists at Karolinska University Hospital in Stockholm who have undergone examinations such as bronchoalveolar lavage. Data is therefore available only for a subset of all sarcoidosis cases identified through the NPR and include disease phenotype (Löfgren's and non-Löfgren's disease), smoking status, HLA typing, and results of other examinations. These individuals provided informed consent for their data to be recorded and used in research investigations.

4.5 STUDY DESIGN AND STATISTICAL ANALYSES

Six individual studies are summarized in this thesis. Two were designed to examine risk factors for sarcoidosis and three focused on longer-term outcomes in patients with the disease. The table below shows an overview of the design, exposure, and outcomes used in the individual studies.

Table 9 | Design aspects of the individual studies in this thesis.

Study	Epidemiological design	Study (enrollment) period	Main exposure	Main outcome
I. Familial aggregation of sarcoidosis	Case-control-family study	1964–2013	Sarcoidosis in first or second degree relatives	Sarcoidosis diagnosis in probands
II. Infectious diseases as risk factors for sarcoidosis	Case-control study	2009–2013	Infectious disease diagnosis at least three years before sarcoidosis diagnosis	Sarcoidosis diagnosis
III. Mortality in sarcoidosis	Cohort study	2003–2013	Sarcoidosis diagnosis	All-cause death
IV. Risk of serious infection in sarcoidosis	Cohort study	2003–2013	Sarcoidosis diagnosis	Hospitalization for infectious disease
V. Infection in methotrexate versus azathioprine for sarcoidosis	Target trial emulation	2007–2013	Initiation of methotrexate or azathioprine	Diagnosis of infectious disease within six months
VI. Risk and predictors of heart failure in sarcoidosis	Cohort study	2003–2013	Sarcoidosis diagnosis	Diagnosis of heart failure

4.5.1 Sarcoidosis cases and comparators

Sarcoidosis in the individual studies in this thesis was defined using the NPR. Patients were required to have at least two visits in the NPR's outpatient or inpatient components listing an ICD code for sarcoidosis (ICD-8/9 135, ICD-10 D86). Two ICD-coded visits were considered to be an adequate balance between capturing enough symptomatic cases while minimizing the risk for misclassification of cases due to the inherent difficulty in diagnosing sarcoidosis. This definition has been consistently used to identify several other inflammatory diseases in the NPR. In Study III (mortality in sarcoidosis), one of the first studies in sarcoidosis to be conducted using Swedish registers, a more restrictive definition was used requiring the two visits to be at least 15 days apart and at least one visit should have listed sarcoidosis as the primary discharge diagnosis. A validation study, which was conducted after Study III was completed, showed that a more liberal definition of sarcoidosis (i.e., at least two ICD-coded visits in the NPR) yielded a high positive predictive value of 94% [219]. Therefore, we used this less restrictive definition of sarcoidosis in the subsequent studies included in this thesis.

An exception, however, was Study I, where we used a more liberal definition requiring at least one hospitalization or at least two outpatient visits listing an ICD code for sarcoidosis to identify individuals and their relatives with sarcoidosis. Our ability to identify sarcoidosis in the NPR was limited before the inception of the outpatient component in 2001. Acknowledging potential bias due to sarcoidosis misclassification by using only one hospitalization for sarcoidosis and the fact that hospitalized sarcoidosis was likely more severe than sarcoidosis diagnosed in an outpatient clinic, this definition was used to increase analytical power by detecting more relatives exposed to sarcoidosis.

To compare sarcoidosis to the general population, controls (or comparators) were randomly sampled and individually matched 10:1 to all individuals with sarcoidosis in the data linkage at the time of their diagnosis. Matching variables were year of birth, sex (female or male), and residential location (parish, municipality, or county/region depending on the availability of eligible individuals in each cluster). Comparators with a history of sarcoidosis at the time of selection in the NPR (using the same definition used for sarcoidosis in each study) were excluded.

Acknowledging the difficulty of differentiating sarcoidosis from lung and hematopoietic malignancies, especially when localized in the thoracic cavity, cases and comparators who had such malignancy recorded in the Cancer Register within six months before or after the first ICD-coded sarcoidosis visit or the corresponding period for comparators were excluded. The assumption was that in these cancer patients with a histologically confirmed malignant tumor, the likelihood of a concomitant sarcoidosis diagnosis was small. To further reduce misclassification of sarcoidosis, individuals younger than 18 years and those older than 85 years for whom a diagnosis of sarcoidosis is very infrequent, and the risk of misdiagnosis was potentially high were also excluded.

Highlights in terms of study design and the statistical methods used to analyze the obtained data is described for each individual study below. This section is not comprehensive, hence the reader is encouraged to refer to each publication for more detailed information about each study.

4.5.2 Study I: Familial aggregation of sarcoidosis

We conducted a case-control-family study using cases and controls from the sarcoidosis linkage. The addition of the word “family” to the case-control design denotes the fact that family members of index individuals were unbiasedly selected and directly ascertained in this study. In contrast, relatives were identified through questionnaires or interviews with probands in previous studies, which increased the risk of differential misclassification due to recall bias favoring the cases. Index individuals with and without sarcoidosis whose first and second degree relatives were identified using the Multi-Generation Register and ascertained for sarcoidosis (the exposure) are called probands.

Exposure was the diagnosis of sarcoidosis in first or second degree relatives of proband cases and controls. Assuming effects of genetics are present throughout a proband’s life, no restriction to the timing of relative’s diagnosis with respect to the proband’s diagnosis was enforced in main analyses. The likelihood of sarcoidosis diagnosis in relatives of cases compared to relatives of controls was estimated using conditional logistic regression accounting for the bias introduced by the outcome-based sampling of controls and the matching on age, sex, and residential location. Because sarcoidosis is more likely to be diagnosed in a larger family than in a smaller one, the average number of relatives per case or control was compared to ensure that familial risks were not influenced by an imbalance in family structures among cases and controls. No adjustment of the logistic regression models for any covariates were deemed necessary.

The exposure was parametrized in two ways. In the first, each relative contributed one observation in the analyses and family clusters within the case-control strata were formed. Lack of independence due to familial clustering in this type of analysis was accounted for by using robust sandwich variance estimators. In the second, a joint exposure status was created for each case and control by ascertaining who of the probands had at least one or at least two family members diagnosed with sarcoidosis. To check for effect measure modification, the analyses were further stratified by proband’s age at diagnosis or matching, sex of the proband and relative, and Löfgren’s status (the latter in a subset of probands registered in the Karolinska Clinical Cohort).

Using the principles of quantitative genetics and the liability-threshold model as well as two analytical methods, probit variance component analysis and tetrachoric correlations, we sought to express the observed familial risks for sarcoidosis on the heritability scale [112,220]. Heritability is a measure signifying how much of the overall observed risk for a phenotype in a specified population (herein, sarcoidosis) can be explained by additive genetic effects as

opposed to shared and non-shared environmental effects within families [112]. To decompose covariance owing to additive genetic and shared environmental effects in this analysis, we used full and half siblings that have the same genetic makeup of up to 50% and 25%, respectively.

A major potential issue in Study I was misclassification of the sarcoidosis in proband cases and relatives of proband cases and controls. To address differential misclassification of sarcoidosis in probands that would have led to an overestimation of the true association, relatives were required to be diagnosed at least a year before proband cases or controls were diagnosed thus respecting the timing of events in a case-control study (i.e., exposure should precede the outcome). In addition, because the register-based definition for sarcoidosis was at the time not validated, probabilistic bias analysis methods were used to estimate bounds for the familial relative risk by assuming a positive predictive value following a beta distribution and centered around 80%.

4.5.3 Study II: Infectious diseases as risk factors for sarcoidosis

Study II was a case-control study in which sarcoidosis cases and non-sarcoidosis controls were enrolled between 2009 and 2013. The exposure in this study was history of infectious disease in cases and controls. As shown in Figure 7 below, a latency period of three years from ascertainment of exposure status to the first visit of sarcoidosis or the corresponding time for controls was used to minimize reverse causation bias due to long-lasting preclinical sarcoidosis in some cases and a possible delay in sarcoidosis diagnosis when the date is obtained from the NPR.

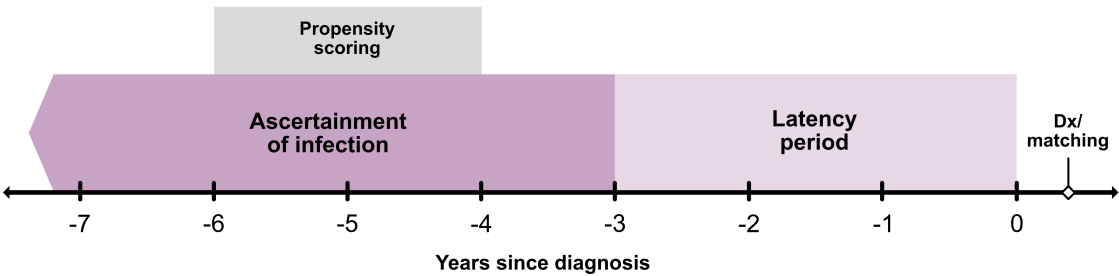


Figure 7 | Graphical representation of the timing of events in Study II.

(Adapted by permission from Springer: *European Journal of Epidemiology*, “Are infectious diseases risk factors for sarcoidosis or a result of reverse causation? Findings from a population-based nested case-control study” by Rossides M, Kullberg S, Askling J, et al. CC BY-NC 4.0, 2020.)

The exposure was parametrized in several different ways to accomplish the various aims in this investigation:

- At least one inpatient or outpatient ICD-coded visit for infectious disease in the NPR was used as primary definition.
- At least one hospitalization for infectious and/or a visit where infectious disease was the primary diagnosis or a combination of both to capture a more severe and likely less misclassified exposure definition.
- At least two inpatient or outpatient ICD-coded visits for infectious disease to capture a higher propensity to infection.
- At least one visit in the NPR or one dispensation of an antimicrobial (i.e., an antibiotic, antiviral, antifungal, or antimycobacterial medication) in the PDR to capture infectious diseases diagnosed in primary care.
- Infectious disease by primary localization, that is, the upper and lower respiratory, gastrointestinal, and genitourinary tracts, the skin, and eyes. Infections in certain organs (e.g., respiratory and ocular) has long been implicated in disease development compared to sites (e.g., the genitourinary tract) that are more rarely affected by sarcoidosis.

The association between different exposures and the risk of developing sarcoidosis in the future was estimated using conditional logistic regression models estimating odds ratios of sarcoidosis and corresponding 95% confidence intervals (CI) accounting for the matching variables birth year, sex, and residential location. Models were further adjusted for deciles of a high-dimensional propensity score [221] for the risk of infectious disease, which was deemed necessary to capture otherwise unmeasured variables that might have resulted in uncontrolled confounding and/or reverse causation bias. The propensity score was estimated using controls' data on visits and medical procedures in the NPR, medication dispensations in the PDR, and included predefined confounders (age at sarcoidosis diagnosis or matching, sex, region of residence, country of birth, education, annual gross salary, civil status, history of autoimmune disease or primary immunodeficiency, number of first degree relatives with history of autoimmune disease, and year of sarcoidosis diagnosis/matching). These data were collected during the period of six to four years before the date corresponding to the first sarcoidosis visit (Figure 7, page 36).

The influence of infectious disease on the risk of developing sarcoidosis was hypothesized to vary by sarcoidosis phenotype. Löfgren's syndrome presents acutely and has better prognosis compared to non-Löfgren's disease [4]. Similarly, treated sarcoidosis around the time of diagnosis is believed to be more severe in terms of symptom presentation and/or organ damage [2], and is probably more likely to be associated with a longer-lasting preclinical phase. The analyses were therefore stratified by sarcoidosis phenotype, that is, Löfgren's versus non-Löfgren's disease using data from the Karolinska Clinical Cohort (available for a subset of cases) and treated compared to untreated sarcoidosis around the time of sarcoidosis diagnosis using the entire study population. Latency periods ranging from zero to seven years from

exposure to outcome ascertainment were also tested to investigate the effect of timing of diagnosis of infectious disease on the risk of developing sarcoidosis.

At study conception, differential misclassification of infectious disease as a result of preclinical sarcoidosis (reverse causation bias) was identified to be a major potential threat to the validity in this study. We took several measures to address reverse causation bias including:

- Imposition of a latency period of at least three years between exposure and outcome ascertainment in the main analyses.
- Use of a high-dimensional propensity score to capture otherwise unmeasured data dimensions that could account for this bias.
- Conduct of probabilistic bias analyses to test the robustness of our findings in the presence of reverse causation bias of varying magnitude.

4.5.4 Study III: Mortality in sarcoidosis

In this cohort study, newly diagnosed patients with sarcoidosis from the NPR and matched general population comparators without sarcoidosis both identified between 2003 and 2013 were followed for all-cause death in the Cause of Death Register. Follow-up started at sarcoidosis diagnosis, which was defined as the second visit for the disease at least 15 days from the first, or the visit for which sarcoidosis was the primary discharge diagnosis, whichever occurred last, or the corresponding date for matched comparators. Follow-up ended at the first of the date of death (outcome of interest), first emigration (data from the Total Population Register), or December 31, 2014 (administrative censoring).

Mortality rates in sarcoidosis and the general population were estimated using Poisson regression models adjusted for age and sex. To compare the sarcoidosis and non-sarcoidosis groups in terms of all-cause death, rate differences and hazard ratios with their corresponding 95% CIs were estimated using Poisson and Cox proportional hazards models, respectively, with time since sarcoidosis diagnosis as the underlying analysis scale. Results from two Cox models adjusted for confounding variables were reported. The first was controlled for the matching variables (i.e., age, sex, and residential location) and the second was further adjusted for country of birth, education, and comorbidity as approximated by the Charlson Comorbidity Index score [222]. Adjusted survival probabilities [223] were estimated using inverse probability weights estimated using the same variables as in the second Cox model. The analyses were stratified by age at inclusion (18–29, 30–39, 40–49, 50–59, 60–69, 70–85 years), sex (female versus male), and sarcoidosis treatment status around the time of diagnosis (± 3 months from the first visit for sarcoidosis in the NPR; treated versus untreated) to examine variations in the hazard ratio of all-cause mortality associated with sarcoidosis.

Smoking was an important confounder of the association between sarcoidosis and all-cause mortality, but no data was available in registers to measure and account for it in the analyses. Using current smoking prevalence estimates from local public health surveys and previous

investigations on the association between smoking, sarcoidosis, and all-cause mortality, we ran simulations (probabilistic bias analyses [209]) to re-estimate the hazard ratio of all-cause death in the presence of unmeasured confounding by current smoking. Moreover, we ran a second probabilistic bias analysis to test the effect of the sarcoidosis definition (which was based on ICD-coded visits that were not validated at the time) on the overall hazard ratio under a scenario of extreme non-differential misclassification (positive predictive value 50–70%). An overall hazard ratio was later obtained by combining the two bias scenarios in a third simulation.

4.5.5 Study IV: Risk of serious infection in sarcoidosis

Study IV was a cohort study in which individuals with sarcoidosis identified from the NPR and matched general population comparators were followed until the diagnosis of the first or separately, several (up to six) serious infections. Serious infection was defined as an inpatient visit in the NPR listing an ICD code for an infectious disease as the primary discharge diagnosis. Participants follow-up was right censored at death from any cause (data from the Cause of Death Register), first emigration since start of follow-up (Total Population Register), or on December 31, 2013, whichever occurred first. A recurrent serious infection was considered a hospital admission for infection that occurred at least 31 days after the previous to minimize the likelihood of considering re-hospitalization for the same infection as a new, unrelated to the previous, infectious disease.

All analytical models were weighted using inverse probability of sarcoidosis weights aiming to estimate marginal rates, risks, and hazard ratios of serious infection comparing sarcoidosis to the general population adjusted for confounding variables. Those included: age, sex, residential location, country of birth, education, gross annual salary, civil status, calendar period, number of visits in the NPR within two years before inclusion, and history of comorbidities (congestive heart disease, atrial fibrillation, hypertension, diabetes mellitus, dyslipidemia, chronic obstructive pulmonary disease, asthma, acute myocardial infarction, stroke, autoimmune disease, and primary immunodeficiency), and history of autoimmune disease or sarcoidosis in at least one first degree relative. To avoid reverse causation bias as individuals with sarcoidosis were more likely to visit healthcare services before diagnosis compared to the general population, history of serious infection in the year before inclusion was not accounted for in the logistic regression model that was used to estimate the weights.

Adjusted incidence rates for serious infection were estimated using weighted Poisson regression models and hazard ratios and corresponding 95% CIs were estimated using weighed Cox proportional hazards regression models with years since study inclusion as the underlying analysis time. Analyses were further stratified by age at start of follow-up (18–44, 45–64, 65–85 years), sex (females versus males), and sarcoidosis treatment status around the time of sarcoidosis diagnosis (± 3 months from the first sarcoidosis visit; treated versus untreated). To examine how rates, risks, and hazard ratios of serious infection varied by follow-up time, flexible parametric survival models were employed [214]. The number and location of knots

used to estimate the underlying hazard of serious infection (on the log cumulative hazard scale) was based on a combination of plotting the hazard functions and minimizing the Akaike Information Criterion of the estimated model. The hazard ratio of recurrent serious infections was estimated using Cox and flexible parametric survival models with a random effects component (a gamma frailty term) to account for the lack of independence among the observed times to infectious disease within an individual. The latter modeling approach was also used to examine whether hazard ratios varied by follow-up time.

In sensitivity analyses, different definitions of serious infection were used to test the robustness of the findings in the presence of misclassification, both non-differential and differential (by sarcoidosis status) resulting in an under- or overestimation of the true association between sarcoidosis and serious infection, respectively. Because individuals with sarcoidosis were on average more likely to be hospitalized than their general population comparators, they were considered to be at higher risk of receiving a diagnosis of an infectious disease. In one analysis, at least one dispensation of an antibiotic, antiviral, antimycobacterial, or antifungal medication within 15 days before or after hospital admission for an infectious disease was required, and in another analysis, pneumonia or urinary tract infection diagnoses were excluded. Moreover, analyses were repeated after excluding cases and comparators with a history of serious infection in the year before inclusion and after restricting the comparator group to individuals who had a history of a healthcare contact (in the NPR) within two years before the start of follow-up for serious infection. Last, the relative risk of serious infection was examined in the subset of sarcoidosis cases diagnosed at Karolinska University Hospital and registered in the local clinical cohort.

4.5.6 Study V: Infection in methotrexate versus azathioprine for sarcoidosis

Study V was an emulation of a target (hypothetical) trial with observational data that aimed to investigate which of methotrexate or azathioprine was associated with a lower risk of infectious disease diagnosis in individuals with sarcoidosis. Methotrexate and azathioprine are the two most popular second line immunosuppressive treatments for sarcoidosis in Sweden [58] and were equally recommended as potential second line choices by sarcoidosis experts [48]. A protocol for the target trial was a priori specified (Table 10, page 41). It was subsequently emulated using information from nationwide Swedish registers, primarily the PDR and the NPR.

Because information on medication dispensations is collected daily in the PDR (day is the unit of time in the register), each day between January 1, 2007 and June 30, 2013 was considered a potential opportunity for initiating a new trial emulation amounting to 2738 potentially successful trial emulations. At each of these 2738 days, eligibility criteria were evaluated. In accordance with the target trial, which was adapted to the observational data at hand, individuals should have fulfilled all of the following five criteria to be considered eligible for inclusion.

They should have:

1. Had at least two ICD-coded inpatient or outpatient visits for sarcoidosis recorded in the NPR up to the day before the trial emulation.
2. Had no hematopoietic or lung malignancy diagnosis recorded in the Cancer Register within six months before or after the first visit for sarcoidosis in the NPR.
3. Been of age 18–85 years at trial emulation.
4. Been dispensed at least one prescription of systemic corticosteroids within six months before trial emulation.
5. Not been dispensed any prescription of either methotrexate or azathioprine within six months before trial emulation.

Table 10 | Protocol for target trial and emulation to investigate infection risks in methotrexate versus azathioprine for sarcoidosis.

Protocol component	Target trial	Trial emulated with observational data
Eligibility criteria	Individuals should be diagnosed with sarcoidosis at any point between 1964 and 2013, be 18 to 85 years old, and live in Sweden at trial entry. They should not have used any of methotrexate or azathioprine in the past six months before trial entry but be current users of systemic corticosteroids. They should also have no active liver or kidney disease	The same inclusion and exclusion criteria will be applied. Initiation of a trial treatment assumes no contraindication to the treatment exists
Treatment strategies	(a) Initiate methotrexate at baseline and remain under study until the follow-up (b) Initiate azathioprine at baseline and remain under study until the follow-up	Same as in the target trial
Assignment procedures	Randomly assigned to initiate each of the treatment strategies. Participants will be aware to which treatment group they belong to	Random assignment will be emulated by accounting for pre-treatment covariates
Follow-up period	Starts at randomization and ends at diagnosis of infectious disease, death, or six months after randomization, whichever occurs first	Same as in the target trial
Outcome	Infectious disease diagnosis at six months from randomization	Same as in the target trial
Causal contrast(s) of interest	Intention-to-treat effect	Same as in the target trial
Analysis plan	Intention-to-treat effect estimated via comparison of six-month risks for infectious disease among individuals assigned to each treatment strategy. All analyses will be adjusted for pre- and post-randomization prognostic factors associated with loss to follow-up (if any)	Same as in the target trial

At each of the potential 2738 trial emulations, if at least one eligible individual was dispensed either methotrexate or azathioprine then they were included in the respective trial arm (methotrexate or azathioprine initiator groups) and subsequently followed for up to six months

for infectious disease (Figure 8). By design, all three of eligibility assessment, initiation of treatment, and start of follow-up coincided at the day of trial initiation. Six months was considered an appropriate time for infectious disease to onset and be diagnosed after initiation of one of the two immunosuppressive treatments [224,225]. Diagnosis of infectious disease was defined as at least one inpatient or outpatient visit in the NPR restricting to those where an ICD code for an infectious disease was allocated as the primary discharge diagnosis. During the study period (January 2007 to June 2013), a unique individual could appear in multiple future trials and/or in different exposure groups (methotrexate or azathioprine) if they were deemed eligible and initiated either methotrexate or azathioprine in future trial emulation instances.

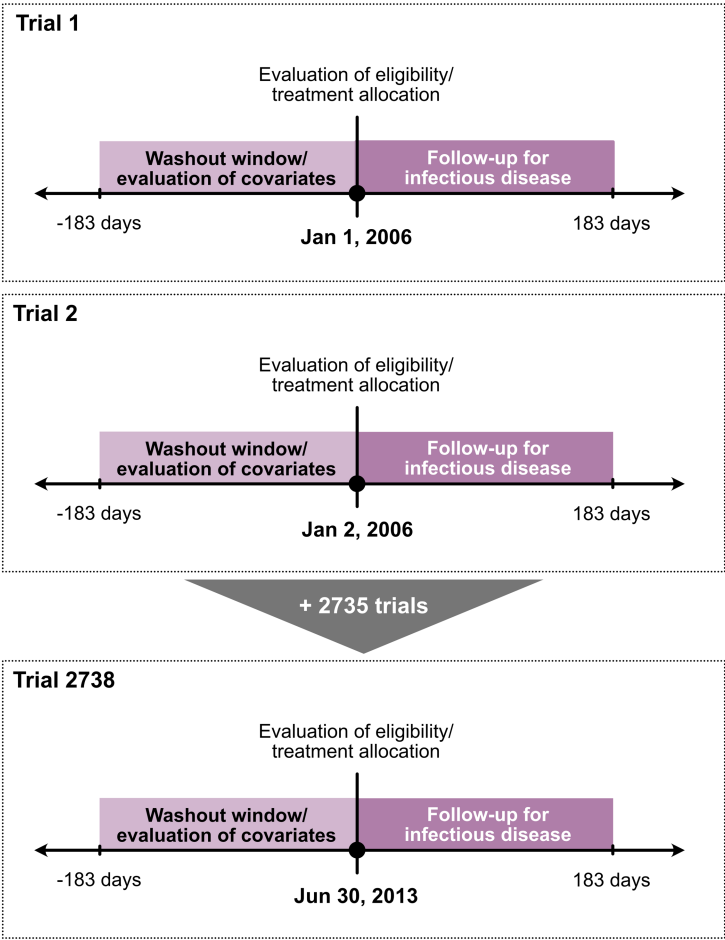


Figure 8 | Schematic representation of the target trial emulation.

A total of 2738 target trial emulations nested in the Prescribed Drug Register could be initiated. At each initiation (i.e., time zero, depicted by a black circle ●), eligibility criteria were evaluated, treatment (methotrexate or azathioprine) was allocated, and the six-month follow-up for infectious disease commenced.

The risk ratio and risk difference at six months after initiation were the targets for estimation of the average treatment effect following an intention-to-treat analysis scheme. Risks, risk ratios, and risk differences were all estimated using targeted maximum likelihood

estimation [216]. Tenfold cross-validation was employed to estimate weighted models of the exposure, outcome, and outcome missingness due to death (a potential source of selection bias) combining conventional statistical modeling approaches with machine-learning algorithms. The same covariates were used to model all three mechanisms: age at trial entry, age at sarcoidosis diagnosis, sex, region of residence, country of birth, education, civil status, calendar period, number of healthcare visits within six months before trial entry, history of comorbidity (congestive heart disease, atrial fibrillation, acute myocardial infarction, stroke, hypertension, diabetes mellitus, dyslipidemia, chronic obstructive pulmonary disease, asthma, and autoimmune disease), dispensation of systemic corticosteroids, non-steroidal anti-inflammatory drugs, or antimicrobials within six months before trial entry. Numbers needed to harm, that is, the average number of patients that need to be treated with methotrexate (compared to azathioprine) for one extra patient to be diagnosed with an infectious disease within six months after treatment start, was calculated using six-month risks estimated from TMLE as $|1/(\text{risk in methotrexate initiators} - \text{risk in azathioprine initiators})|$.

Several sensitivity analyses were conducted to test the robustness of the main analysis to confounding by indication, selection bias, misclassification of the outcome, and choice of the targeted maximum likelihood estimation procedure to estimate the average treatment effect, which was not used before in these data. Specifically, individuals with sarcoidosis were required to never have been dispensed methotrexate or azathioprine to be eligible for inclusion, the follow-up period was shortened to three or extended to nine months, and infectious diseases were defined solely based on hospitalizations, or using dispensations of antimicrobial medications, or deaths due to infectious disease in addition to healthcare visits. All analyses were repeated using modified Poisson regression [226], a widely used method to obtain risk ratios. To estimate how much unmeasured confounding would be needed to explain the observed effect from the main analysis, we calculated the E-value [227]. It reflected the magnitude of an association (on the risk ratio scale) between a potential unmeasured confounding variable with both the exposure (methotrexate versus azathioprine initiation) and the outcome (infectious disease diagnosis within six months after treatment initiation) to move the observed risk ratio to the null value of one.

4.5.7 Study VI: Risk and predictors of heart failure in sarcoidosis

We conducted a cohort study to examine the relative risk of heart failure associated with sarcoidosis and identify clinical predictors of heart failure in sarcoidosis. In this study, individuals with sarcoidosis diagnosed between 2003 and 2013 and matched general population comparators without sarcoidosis were followed for a heart failure diagnosis in the NPR. Heart failure was defined as an inpatient or outpatient visit in the NPR listing an ICD code for heart failure or cardiomyopathy as the primary discharge diagnosis, a definition that yielded a high (>80%) positive predictive value in several validation studies [228]. Follow-up for heart failure started at the second visit for sarcoidosis in the NPR or the corresponding date for matched comparators and ended at the date of first hospital admission

or outpatient visit for heart failure, death (data from the Cause of Death Register), first emigration (Total Population Register), or December 31, 2013, whichever occurred first. Individuals were eligible for follow-up if they had no diagnosis of heart failure at start of follow-up.

Data from administrative and health registers was collected to evaluate several demographic and clinical variables which served as confounders in the analysis for the relative risk of heart failure associated with sarcoidosis and as predictors of heart failure in sarcoidosis. Those included: birth year (to calculate age), sex, region of residence, country of birth, years of completed education, civil status, number of visits within two years before sarcoidosis diagnosis or matching, and comorbidity (hypertension, diabetes mellitus, dyslipidemia, ischemic heart disease including acute myocardial infarction, heart valve disease, atrial fibrillation, other [ventricular] arrhythmias including heart blocks, chronic obstructive pulmonary disease, chronic kidney disease, alcohol-related disorders, and autoimmune disease).

Incidence rates of heart failure were estimated using Poisson regression models adjusted for age at start of follow-up, sex, and region of residence. To estimate the hazard ratios of heart failure comparing sarcoidosis to the general population, Cox proportional hazards models with attained age as the underlying time scale were used. They were progressively adjusted for (1) the matching variables age, sex, and region of residence, (2) other demographic variables (country of birth, education, and civil status) and calendar period, and (3) morbidity that was evaluated three months before the first visit for sarcoidosis in the NPR or the corresponding date for comparators (number of visits in the NPR in the past two years, hypertension, diabetes mellitus, dyslipidemia, heart valve disease, atrial fibrillation, chronic obstructive pulmonary disease, chronic kidney disease, alcohol-related disorders, and autoimmune disease). The proportional hazards assumption was found to hold when Schoenfeld residuals plots were examined.

The analyses were further stratified by pre-determined effect measure modifiers: age at start of follow-up (18–44, 45–64, or 65–85 years), sex (female versus male), treatment with an immunosuppressant around the time of sarcoidosis diagnosis (treated versus untreated), time since sarcoidosis diagnosis (≤ 2 versus > 2 years after start of follow-up), and history of ischemic heart disease or acute myocardial infarction at start of follow-up (yes versus no). Sarcoidosis is expected to enter remission within two years after diagnosis and ischemic heart disease is a major cause of ischemic heart failure [2,187]. Statistically significant effect measure modification was indicated by a *P*-value smaller than 0.05 from a likelihood ratio test.

To examine predictors of heart failure in sarcoidosis, the analytical sample was restricted to individuals with sarcoidosis. A multivariable Cox proportional hazards model with years since sarcoidosis diagnosis as the analysis time scale was used to determine which clinical predictors were independently associated with a higher risk of heart failure in sarcoidosis. Covariates included in the model were demographics (age at sarcoidosis diagnosis, sex, region of residence, country of birth, education) and morbidity evaluated at start of follow-up (hypertension, diabetes mellitus, ischemic heart disease including acute myocardial infarction,

heart valve disease, atrial fibrillation, other arrhythmias including heart block and non-fatal cardiac arrest, chronic kidney disease, chronic obstructive pulmonary disease, alcohol-related disorders, and autoimmune disease). To rank the clinical predictors by their relative contribution to heart failure cases in sarcoidosis, the attributable fraction at two and 10 years was estimated considering the strength of the association between each predictor and heart failure and their prevalence at start of follow-up [229].

Several sensitivity analyses were conducted to test the robustness of the main findings. We disregarded heart failure diagnoses during the first six months of follow-up in sarcoidosis and the general population to account for potential differential misclassification of heart failure due to surveillance bias favoring sarcoidosis. We also examined another definition of heart failure including secondary diagnoses in the NPR in addition to primary discharge diagnoses. For our analysis of heart failure predictors in sarcoidosis, time-varying clinical predictors (morbidity) were also examined. Last, immunosuppressant treatment was considered as an additional predictor of heart failure in sarcoidosis. It was defined as treatment status around diagnosis (treated versus untreated), or as the cumulative defined daily doses of systemic corticosteroids that were dispensed to an individual within six months before or, separately, six months after sarcoidosis diagnosis. All analyses for sarcoidosis treatment were restricted to patients with sarcoidosis diagnosed starting 2006 and onwards for whom data on filled prescriptions was available in the PDR.

4.6 ETHICAL PERSPECTIVES

Ethical research should possess value, be timely, valid, cost-effective and beneficial to individuals and/or the society. No research is harmless, however, but when benefits surpass potential risks, a research project is ethically permissible. In 2014, the Regional Ethics Review Board in Stockholm, which according to Swedish law oversaw at the time the ethical aspects of research on human subjects, granted ethical approval for the register linkage and the projects included in this thesis (protocol number 2014/230-31).

The individual studies and my doctoral project as a whole created value for current and future patients, physicians, researchers, and the society. The scope of this project was multifold. A major aim was to aid diagnosis and treatment of current and future patients with sarcoidosis and reduce the burden of disease in patients and the society. Another aim was to set priorities and guide future research efforts. Considering the scarcity of research in this field, another goal was to provide as much research output as possible while keeping costs to the minimum. The added value of this project exceeded both planned and unexpected costs by using already collected data and methods to maximize inference.

The research questions were valid and timely at the time of conduct and were answered using appropriate study designs and analytical methods. To the project's advantage, the use of population-based high-quality registers allowed for both power and bias mitigation. It should be noted, however, that register-derived data and the methods used to analyze those are far

from perfect. Imperfections in study design and data analyses were addressed with sensitivity or bias analyses designed to set bounds to those errors that might have affected the interpretation of study results.

No direct contact with research subjects included in the studies (individuals with sarcoidosis, comparators, or their relatives) was established during the conduct of this doctoral project. Nevertheless, access to de-identified personal information that exists in administrative and health databases was required to complete the individual studies. The mere existence of those databases poses risks for the individual: violation of autonomy, privacy, and personal integrity that may inflict stigma and economic consequences. However, as reflected in the local law, the Swedish society accepts the access to and the processing of those data for research purposes under certain conditions and in the presence of ethical permission. To eliminate any risk of bridging participants' confidentiality, rules and policies governing the possession and processing of personal data were strictly followed: all data was de-identified, encrypted, and securely stored in dedicated on-premise servers and access was limited to those directly involved in the work with security clearance to do so.

Benefits of the conducted studies exceeded possible harms. Potential benefits were small for study participants, especially from research on disease etiology, but extensive for future patients and the society. Studies concerned with the etiology of sarcoidosis are unlikely to confer any direct or indirect benefits to current patients, although studies on patient outcomes have and are expected to indirectly benefit current patients by increasing knowledge, awareness, and practice among physicians and patients. In terms of risks, breach of confidentiality is the only identifiable harm for subjects participating in the studies included in this thesis. However, the likelihood of this happening was and remains minimal because of the strict measures outlined above that were put in place to ensure unidentifiability of individual participants at all stages of the research process.

5 RESULTS

5.1 THE SARCOIDOSIS GROUP AT DIAGNOSIS

Individuals with sarcoidosis were on average 50 years old at sarcoidosis diagnosis (second visit for sarcoidosis in the NPR). Females made up 45% of the sarcoidosis group. The majority of cases (approximately 90%) were born in a Nordic country (i.e., in Sweden, Norway, Denmark, Finland, or Iceland) and about one in three completed university-level education and declared an annual gross salary close to the median salary in Sweden during the study period (360 000 Swedish krona) [230]. Half of the patient population were registered as either being married or living with a partner at sarcoidosis diagnosis. ICD-coded visits used to identify these patients in the NPR indicated that about 80% of patients were diagnosed in an outpatient pulmonary or internal medicine clinic [58].

Table 11 | Comorbidity and medication use at sarcoidosis diagnosis or matching.

	Sarcoidosis	General population
<i>Comorbidity^a</i>	(n=8737)	(n=86 376)
Congestive heart disease	2.4%	1.3%
Atrial fibrillation	3.2%	2.1%
Acute myocardial infarction	2.1%	1.8%
Stroke	1.7%	1.6%
Chronic obstructive pulmonary disease	2.3%	1.0%
Asthma	4.6%	2.4%
Hypertension	21.4%	15.9%
Diabetes mellitus	7.5%	4.2%
Dyslipidemia	10.8%	8.2%
Autoimmune disease	7.9%	4.3%
Primary immunodeficiency	0.4%	0.1%
<i>Medication use within six months before diagnosis/matching^b</i>	(n=6723)	(n=66 441)
Systemic corticosteroids	18.7%	2.9%
Other immunosuppressants ^c	1.2%	0.7%
Inhaled corticosteroids	7.3%	1.9%
Non-steroidal anti-inflammatory drugs	26.3%	9.7%
Antimicrobials ^d	32.6%	13.2%

^aComorbidities were defined using healthcare visits in the National Patient Register and medication dispensations in the Prescribed Drug Register, if applicable.

^bMedication use was defined as at least one dispensation within six months before sarcoidosis diagnosis or matching in individuals included in the cohort starting January 1, 2006 for whom data in the Prescribed Drug Register was available.

^cOther immunosuppressants include methotrexate, azathioprine, and leflunomide.

^dAntimicrobials include antibacterial, antimycobacterial, antifungal, and antiviral medications.

At sarcoidosis diagnosis, comorbid conditions were more prevalent in sarcoidosis cases compared to the general population. Table 11 (page 47) provides an overview of the distribution of relevant comorbidities and medication use in sarcoidosis cases and general population comparators matched for age, sex, and residential location who were identified between 2003 and 2013. As previously shown [58], many individuals with sarcoidosis were in contact with healthcare for about three to six months before sarcoidosis was diagnosed in the NPR (see Figure 5, page 10). Approximately 40% of individuals were dispensed at least one prescription of systemic (oral) corticosteroids, methotrexate, or azathioprine within three months before or after the first visit for sarcoidosis in the NPR. These patients were considered treated around the time of sarcoidosis diagnosis, representing a group with more severe disease. Prescription dispensation data in the PDR was available for all individuals with sarcoidosis diagnosed starting 2006, that is, about 75% of the newly diagnosed sarcoidosis population enrolled between the years 2003 and 2013.

5.2 MAIN FINDINGS BY INDIVIDUAL STUDY

5.2.1 Study I: Familial aggregation of sarcoidosis

A total of 23 880 proband cases and 171 891 proband controls with at least one first or second degree relative were included in this study. Around four first degree relatives per proband cases or control were identified with similar age and sex distributions between the two groups. Comparisons of familial risks between cases and controls were therefore not impacted by imbalances in the identifiability of relatives of proband cases and controls or the ascertainment of sarcoidosis in those.

Table 12 | Familial relative risks of sarcoidosis associated with having relatives with the disease.

Exposure	Exposed/ Total cases	Exposed/ Total controls	Familial relative risk of sarcoidosis (95% CI)
≥1 first degree relative with sarcoidosis			
Overall	831/20 332	1907/164 628	3.73 (3.43, 4.06)
Age of proband 18–49 years	438/10 138	975/85 445	3.99 (3.55, 4.48)
Age of proband ≥50 years	393/10 184	932/79 183	3.48 (3.08, 3.92)
Proband with Löfgren's syndrome ^a	15/356	31/2812	4.14 (2.21, 7.75)
Proband with non-Löfgren's sarcoidosis ^a	21/627	49/4843	3.32 (1.98, 5.56)
≥2 first degree relatives with sarcoidosis	28/20 332	49/164 628	4.69 (2.93, 7.51)
Half sibling with sarcoidosis	44/7511	49/15 202	1.50 (0.98, 2.30)

CI = confidence interval.

^aData on Löfgren's syndrome and non-Löfgren's disease was obtained in a subset of probands registered in the Karolinska Clinical Cohort.

The main results are summarized in Table 12 (page 48). Four percent of proband sarcoidosis cases had at least one first degree relative diagnosed with the disease, which was associated with an almost fourfold increased risk for developing the disease (familial relative risk 3.73 [95% CI 3.43, 4.06]). Increasing the number of exposed first degree relatives to at least two led to an increased familial relative risk of 4.69 (95% CI 2.93, 7.51), whereas considering only half siblings resulted in an attenuation of the familial relative risk for sarcoidosis to 1.50 (95% CI 0.98, 2.30).

Stratifying the analyses by age of the proband at the mean age at sarcoidosis diagnosis (50 years), we observed a somewhat higher relative risk in those diagnosed at younger age (3.99 versus 3.48). No differences were identified in analyses stratified by sex of the proband and relative (data not shown). Despite small numbers, an analysis by sarcoidosis phenotype (Löfgren’s versus non-Löfgren’s disease) in the Karolinska Clinical Cohort indicated a higher familial relative risk for probands who received a diagnosis of Löfgren’s than non-Löfgren’s sarcoidosis (4.14 versus 3.32).

The familial relative risk did not materially change in the presence of potential non-differential misclassification of sarcoidosis in probands and their relatives and differential misclassification due to the unrestricted timing of sarcoidosis diagnosis between relatives and probands in the main analysis, which might have led to reverse causality.

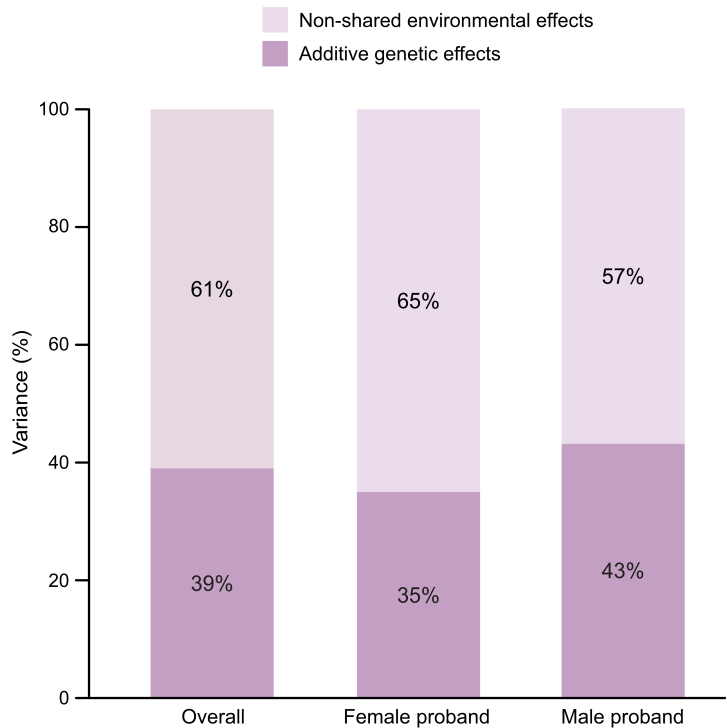


Figure 9 | Heritability of sarcoidosis.

Additive genetic effects (heritability) and non-shared environmental effects contributing to sarcoidosis risk in the Swedish population, overall and by sex of the proband.

(Adapted by permission from the European Respiratory Society: *European Respiratory Journal*, “Familial aggregation and heritability of sarcoidosis: a Swedish nested case-control study” by Rossides M, Grunewald J, Eklund A, et al. Copyright, 2018.)

Using data from full and half siblings, 39% of the susceptibility to sarcoidosis (95% CI 12%, 65%) was found to be attributed to additive genetic effects (heritability) in a probit variance decomposition model with additive genetic and non-shared environmental effects (Figure 9, page 49). No within-family shared environmental effects could be identified and the heritability did not differ between female or male probands. Using tetrachoric correlations and the prevalence of sarcoidosis in our population to estimate heritability yielded an overall similar result (35%).

5.2.2 Study II: Infectious diseases as risk factors for sarcoidosis

In this study, we included 4075 individuals with sarcoidosis (cases) and 40 688 non-sarcoidosis general population controls diagnosed between 2009 and 2013. Twenty one percent of cases and 16% of controls had at least one visit for any infectious disease three years before the first visit for sarcoidosis or the corresponding date for controls. Having at least one visit for infectious disease three years before sarcoidosis diagnosis (or matching) was associated with an odds ratio of sarcoidosis of 1.19 (95% CI 1.09, 1.29) after adjusting for several confounders through a high-dimensional propensity score (Table 13). Odds ratios of similar magnitude were observed for other parametrizations of infectious disease requiring at least two healthcare visits, restricting to hospitalizations for infectious disease, or including antimicrobial dispensations in addition to visits in the NPR.

Table 13 | Relative risk of sarcoidosis associated with history of infectious disease.

Odds ratio of sarcoidosis associated with history of infectious disease, overall and by site of infection. A three-year latency period between infectious disease ascertainment and sarcoidosis diagnosis was applied.

Exposure definition ^a	History of infectious disease diagnosis		Odds ratio of sarcoidosis ^b (95% CI)
	Sarcoidosis (n=4075)	General population (n=40 688)	
≥1 visit for infectious disease	846 (20.8)	6461 (15.9)	1.19 (1.09, 1.29)
≥2 visits for infectious disease	372 (9.1)	2520 (6.2)	1.28 (1.14, 1.44)
≥1 visit for infectious disease or ≥1 dispensation of an antimicrobial	2260 (55.5)	19 589 (48.1)	1.21 (1.13, 1.30)
≥1 visit for respiratory infection or ≥1 dispensation of an antimycobacterial or an influenza antiviral medication	280 (6.9)	1878 (4.6)	1.25 (1.10, 1.42)
≥1 visit for skin infection or ≥1 dispensation of an acne or a herpes zoster antiviral medication	319 (7.8)	2580 (6.3)	1.15 (1.02, 1.29)
≥1 visit for ocular infection	12 (0.3)	72 (0.2)	1.93 (1.12, 3.33)
≥1 visit for gastrointestinal infection	120 (2.9)	940 (2.3)	1.15 (0.95, 1.39)
≥1 visit for genitourinary infection	267 (6.6)	2202 (5.4)	1.06 (0.92, 1.21)

CI = confidence interval.

Data are n (%) unless otherwise stated.

^aVisits refer to hospitalizations or outpatient visits to specialists recorded in the National Patient Register.

^bEstimated using conditional logistic regression models adjusted for deciles of a high-dimensional propensity score.

When sites of infection were considered (Table 13, page 50), ocular infectious diseases (that occurred at least three years before sarcoidosis) were associated with an almost twofold increased odds for developing sarcoidosis, although these infections were observed in very few cases or controls. Having a history of respiratory infection at least three years before sarcoidosis diagnosis was associated with 25% increased odds of sarcoidosis (odds ratio 1.25 [95% CI 1.10, 1.42]), whereas a history of skin, gastrointestinal, or genitourinary infections was weakly associated (if at all) with sarcoidosis diagnosis in the future.

A stronger association between infectious disease history and sarcoidosis was found for cases who were treated for sarcoidosis around diagnosis than those who were not treated (odds ratio 1.41 versus 1.09, respectively). As shown in Figure 10 below, compared to the odds ratio from the main analysis (1.2) where a latency period of three years between infectious disease ascertainment and sarcoidosis diagnosis was required, no differences were observed in odds ratios when lag times of one to seven years were tested. Requiring no lag time resulted in an increase of the odds ratio to 1.5.

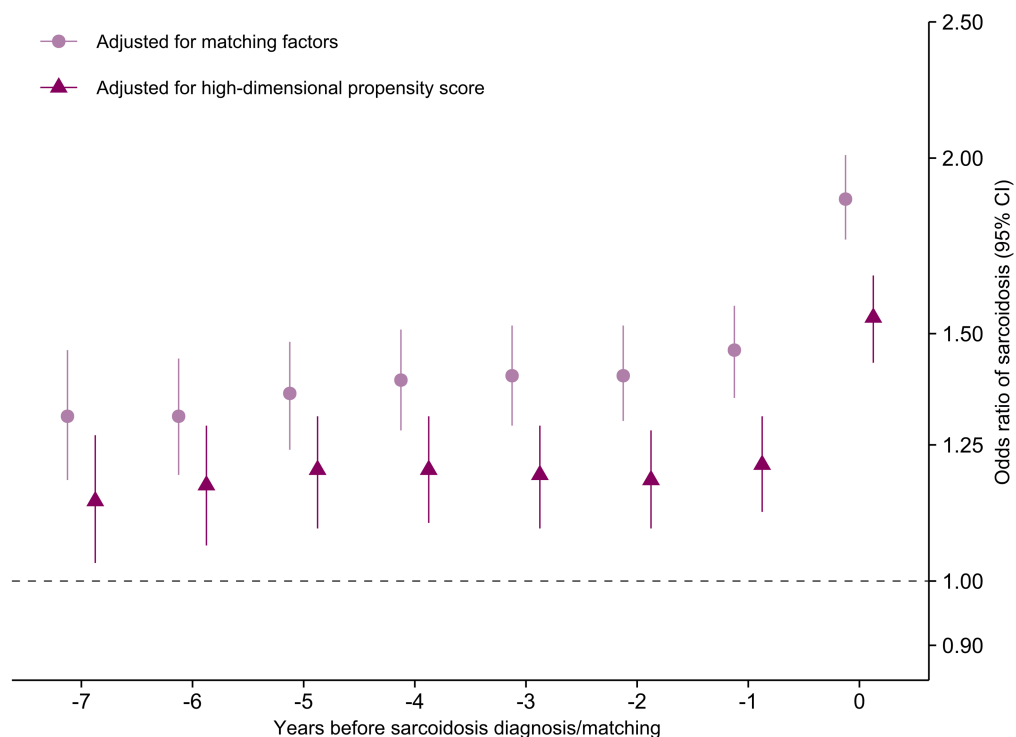


Figure 10 | Odds ratio of sarcoidosis associated with a history of infectious disease by latency period. Latency period refers to the time between ascertainment of history of infectious disease and the first visit for sarcoidosis in the National Patient Register or the corresponding date in matched controls. (CI = confidence interval)

(Adapted by permission from Springer: *European Journal of Epidemiology*, “Are infectious diseases risk factors for sarcoidosis or a result of reverse causation? Findings from a population-based nested case-control study” by Rossides M, Kullberg S, Askling J, et al. CC BY-NC 4.0, 2020.)

The association between history of infectious disease and sarcoidosis was notably attenuated in probabilistic bias analyses designed to simulate different scenarios of differential misclassification of the exposure (reverse causation bias). A null odds ratio of sarcoidosis was estimated when at least one in 10 sarcoidosis cases were assumed to have developed an infectious disease because of long-standing preclinical sarcoidosis (bias adjusted odds ratio 1.02 [95% simulation interval 0.90, 1.15]).

5.2.3 Study III: Mortality in sarcoidosis

A total of 8207 individuals with sarcoidosis and 81 119 non-sarcoidosis comparators were followed for all-cause death. At inclusion, sarcoidosis cases were more likely to be diagnosed with comorbid conditions compared to comparators (mean Charlson Comorbidity Index score 0.24 versus 0.13, respectively).

Table 14 | Risk of all-cause death in sarcoidosis compared to the general population.
Adjusted hazard ratios of all-cause death comparing sarcoidosis to the general population, overall and stratified by age at inclusion, sex, and sarcoidosis treatment status around diagnosis.

	Sarcoidosis	General population	Hazard ratio^a (95% CI)
Overall sarcoidosis	528/8207 (11.0)	3204/81 119 (6.7)	1.61 (1.47, 1.76)
Age at start of follow-up, years			
18–29	2/630 (0.9)	24/6307 (0.5)	0.69 (0.16, 2.96)
30–39	25/1911 (1.6)	109/18 809 (0.9)	1.62 (1.02, 2.56)
40–49	41/1907 (3.0)	192/18 918 (1.8)	2.03 (1.45, 2.85)
50–59	80/1655 (8.6)	476/16 409 (5.0)	1.54 (1.21, 1.96)
60–69	156/1303 (21.0)	838/12 837 (12.3)	1.65 (1.38, 1.96)
70–85	224/801 (72.5)	1565/7 839 (42.5)	1.52 (1.32, 1.75)
Sex			
Female	260/3613 (11.5)	1659/35 765 (6.7)	1.55 (1.36, 1.77)
Male	268/4594 (15.2)	1545/45 354 (8.9)	1.68 (1.47, 1.91)
Sarcoidosis treatment around diagnosis ^b			
Treated	183/2599 (15.4)	826/25 726 (6.8)	2.34 (1.99, 2.75)
Untreated	136/3592 (8.3)	1033/35 491 (6.2)	1.13 (0.94, 1.35)

CI = confidence interval.

Data are number of deaths/number of individuals at risk (age- and sex-adjusted mortality rate per 1000 person-years) unless otherwise stated.

^aEstimated using Cox proportional hazards models with time since sarcoidosis diagnosis or matching as the underlying time scale and adjusted for age, sex, residential location, country of birth, education, and comorbidity (Charlson Comorbidity Index score).

^bEvaluated in individuals with sarcoidosis and their matched general population comparators who entered the cohort in October 1, 2005 and onwards for whom prescription dispensation data was available in the Prescribed Drug Register (established in July 2005).

After a median follow-up of 5.9 years (interquartile range 3.4, 8.7 years), 528 deaths were identified in the sarcoidosis group and 3204 in the comparators group. The all-cause mortality rate was 11.0 per 1000 person-years (95% CI 10.1, 12.0) in sarcoidosis compared to 6.7 per 1000 person-years (95% CI 6.5, 6.9) in the general population (Table 14, page 52). After adjusting for demographic variables and comorbidity, sarcoidosis was associated with a 61% increased risk of all-cause death compared to the general population (hazard ratio 1.61 [95% CI 1.47, 1.76]). The hazard ratio remained largely robust in probabilistic bias analyses aiming to simulate scenarios of unmeasured confounding by current smoking and non-differential misclassification of the sarcoidosis definition (overall hazard ratio 1.66 [95% simulation interval 1.40, 1.93]).

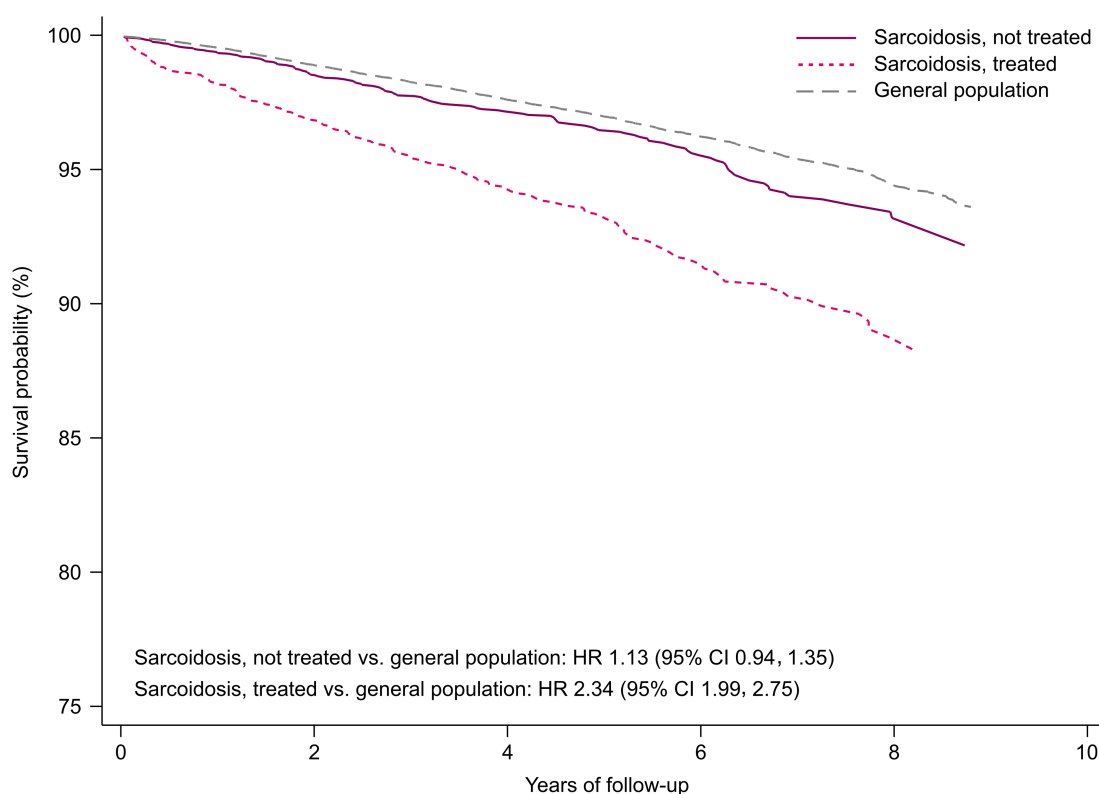


Figure 11 | Survival in sarcoidosis by treatment status.

Adjusted survival probabilities in sarcoidosis and the general population stratified by sarcoidosis treatment status around the time of diagnosis.

(HR = hazard ratio; CI = confidence interval)

(Adapted by permission from the European Respiratory Society: *European Respiratory Journal*, “Sarcoidosis mortality in Sweden: a population-based cohort study” by Rossides M, Kullberg S, Askling J, et al. CC BY-NC 4.0, 2018.)

No notable differences were identified in the hazard ratio of all-cause mortality in stratified analyses by age or sex. As depicted in Figure 11 above, marked variation was observed when analyses were stratified by treatment status around diagnosis: treated individuals had a 2.3-fold higher risk of all-cause death compared to the general population (hazard ratio 2.34 [95% CI

1.99, 2.75]), while the risk for untreated sarcoidosis cases was not markedly different to that of their general population comparators (hazard ratio 1.13 [95% CI 0.94, 1.35]).

5.2.4 Study IV: Risk of serious infection in sarcoidosis

A total of 8737 newly diagnosed sarcoidosis cases and 86 376 non-sarcoidosis comparators from the general population were followed for first and recurrent diagnoses of serious infections in the NPR. Table 15 summarizes the main findings of Study IV.

During a median follow-up time of five years, 895 individuals with sarcoidosis and 3881 comparators developed a first serious infection. The adjusted incidence rate of serious infection was 17.4 per 1000 person-years (95% CI 16.0, 18.9) in the sarcoidosis group and 9.6 per 1000 person-years (95% CI 9.3, 9.9) in comparators. Risks by follow-up time are presented in Figure 12 (page 55). After adjusting for confounders, sarcoidosis was associated with a 1.8-fold increased risk of first serious infection compared to the general population (hazard ratio 1.81 [95% CI 1.65, 1.98]).

Table 15 | Risk of serious infection comparing sarcoidosis to the general population.

Adjusted hazard ratios of first serious infection comparing sarcoidosis to the general population, overall and stratified by age at start of follow-up, sex, and sarcoidosis treatment status around diagnosis.

	Sarcoidosis	General population	Hazard ratio ^a (95% CI)
Overall sarcoidosis	895/8737 (17.4)	3881/86 376 (9.6)	1.81 (1.65, 1.98)
Age at start of follow-up, years			
18–44	222/3712 (8.8)	937/36 698 (5.1)	1.74 (1.44, 2.09)
45–64	327/3447 (16.1)	1334/34 269 (8.5)	1.90 (1.64, 2.20)
65–85	346/1578 (58.3)	1610/15 409 (27.0)	2.16 (1.88, 2.49)
Sex			
Female	435/3890 (21.6)	1938/38 505 (10.7)	2.01 (1.78, 2.28)
Male	460/4847 (14.3)	1943/47 871 (8.7)	1.64 (1.45, 1.87)
Sarcoidosis treatment around diagnosis ^b			
Treated	326/2762 (29.8)	948/27 325 (9.8)	3.04 (2.61, 3.55)
Untreated	275/3961 (15.7)	1428/39 116 (10.2)	1.53 (1.31, 1.80)

CI = confidence interval.

Data are number of first serious infections after start of follow-up/number of individuals at risk (adjusted incidence rate per 1000 person-years estimated using Poisson models weighted for inverse probability of sarcoidosis weights) unless otherwise stated.

^aEstimated using Cox proportional hazard models with time since sarcoidosis diagnosis or matching as the time scale and weighted for inverse probability of sarcoidosis weights.

^bEvaluated in individuals with sarcoidosis and their matched general population comparators who entered the cohort in 2006 and onwards for whom prescription dispensation data was available in the Prescribed Drug Register (established in July 2005).

In stratified analyses, the hazard ratio of serious infection was higher in females and in individuals who received sarcoidosis treatment around the time of diagnosis compared to males and those who were not treated, respectively (Table 15, page 54). Considerable variation in the hazard ratio of serious infection was observed by years since sarcoidosis diagnosis (or the corresponding time for comparators) with a threefold increased relative risk around inclusion leveling off to 1.4 two years after start of follow-up (Figure 13, page 56).

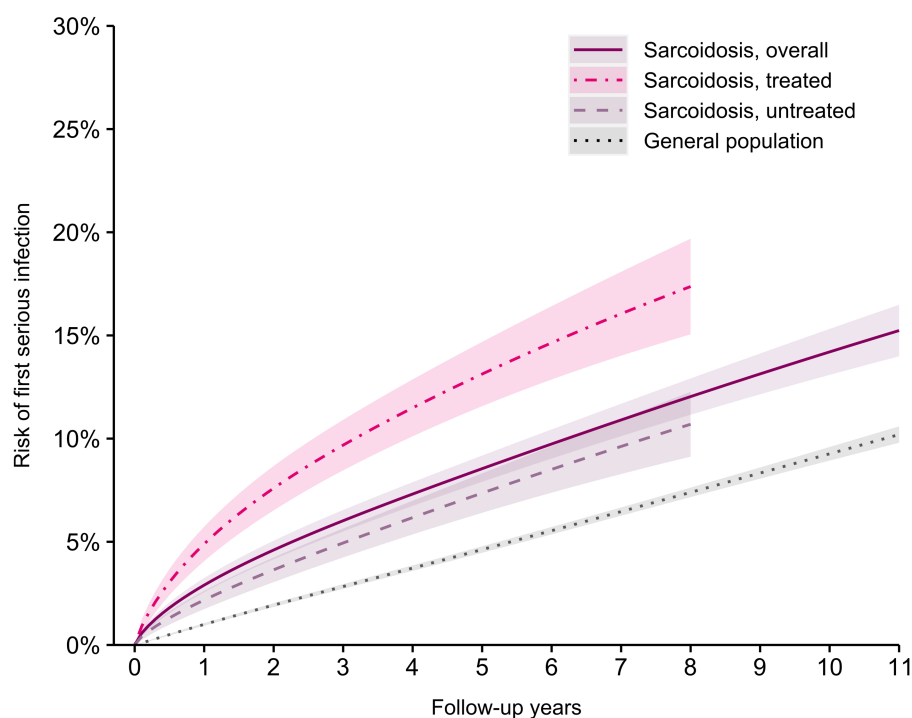


Figure 12 | Adjusted risk of first serious infection by years since sarcoidosis diagnosis or matching.

Follow-up for treated and untreated sarcoidosis was shorter as data on filled prescriptions in the Prescribed Drug Register was available for individuals diagnosed starting in January 1, 2006.

(Adapted by permission from the European Respiratory Society: *European Respiratory Journal*, “Risk of first and recurrent serious infection in sarcoidosis: a Swedish register-based cohort study” by Rossides M, Kullberg S, Eklund A, et al. CC BY-NC 4.0, 2020.)

Higher relative risks of first serious infection associated with sarcoidosis were consistently observed in sensitivity analyses in which antimicrobial dispensations were required in addition to hospitalizations for infectious disease, when pneumonia or urinary tract infections were excluded, and in an analysis restricted to individuals registered in the Karolinska Clinical Cohort. However, the relative risk of serious infection for sarcoidosis overall and especially for not treated sarcoidosis was attenuated when the comparator population was restricted to individuals who had a history of at least one visit in the NPR within two years before they were matched to sarcoidosis cases aiming to obtain individuals who were likely more prone to hospitalization or to have a history of morbidity (hazard ratio for overall sarcoidosis 1.23 [95% CI 1.12, 1.35]; for untreated sarcoidosis 1.03 [95% CI 0.88, 1.22]).

Individuals with sarcoidosis were more likely to be diagnosed with more than one consecutive serious infections than comparators (5.6% versus 1.8%). The adjusted hazard ratio of serious infection recurrence comparing sarcoidosis to the general population was 2.79 (95% CI 2.51, 3.10).

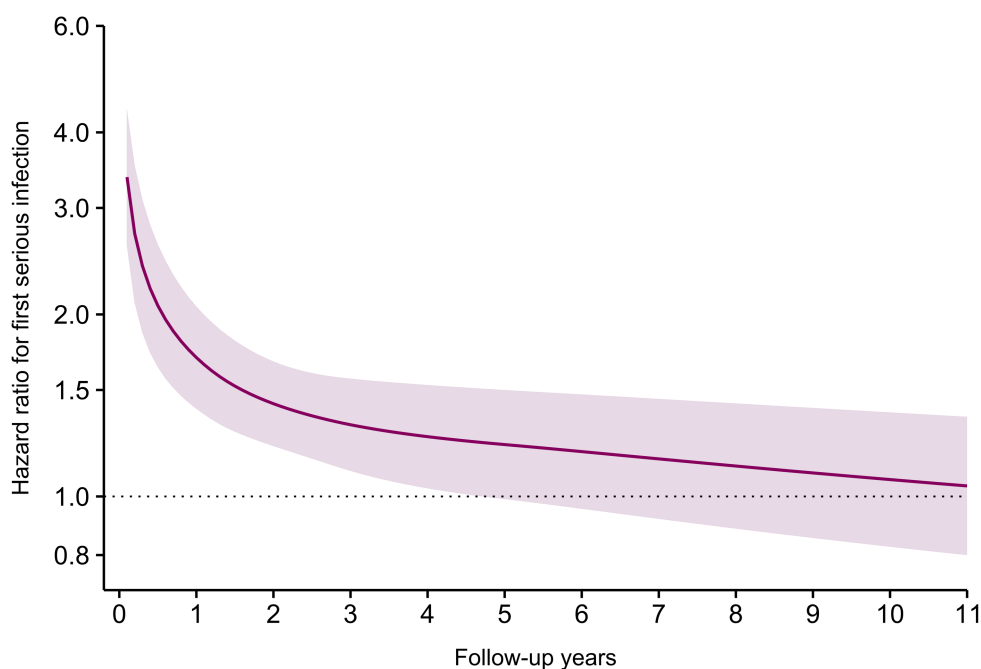


Figure 13 | Adjusted hazard ratio of first serious infection by years since sarcoidosis diagnosis comparing sarcoidosis to the general population.

(Adapted by permission from the European Respiratory Society: *European Respiratory Journal*, “Risk of first and recurrent serious infection in sarcoidosis: a Swedish register-based cohort study” by Rossides M, Kullberg S, Eklund A, et al. CC BY-NC 4.0, 2020.)

5.2.5 Study V: Infection in methotrexate versus azathioprine for sarcoidosis

Of the 2738 potential trial emulations conducted between January 1, 2007 and June 30, 2013, 667 methotrexate and 259 azathioprine initiation episodes were identified. A total of 493 unique individuals initiated methotrexate and 231 azathioprine. Of these, very few (<6%) initiated methotrexate while having a history of azathioprine use or the opposite. In both groups, initiation of the second line treatment occurred within a median time of three years from sarcoidosis diagnosis. Initiators of methotrexate, however, were dispensed a somewhat lower dose of systemic corticosteroids (mean 196 defined daily doses) within six months before initiation of methotrexate compared to azathioprine initiators (mean 238 defined daily doses of systemic corticosteroids).

Within six months after initiation of methotrexate, 43 infectious disease diagnoses were observed in the methotrexate group compared to 29 in the azathioprine group. Death prevented ascertainment of infectious disease at six months in six methotrexate and six azathioprine

initiation episodes. The adjusted six-month risk for infectious disease was estimated to be 6.8% for methotrexate (95% CI 5.3%, 8.6%) and 12.0% in azathioprine initiators (95% CI 10.0%, 14.3%). Compared to azathioprine, methotrexate initiation was associated with a 43% lower risk of infectious disease diagnosis at six months (six-month risk ratio 0.57 [95% CI 0.39, 0.82]). In absolute terms, a reduction of approximately five percentage points in the risk of infectious disease was observed after initiation of methotrexate compared to azathioprine (risk difference -5.17 [95% CI -8.53%, -1.82%]). On average, 19 individuals needed to initiate methotrexate compared to azathioprine for one to be diagnosed with an infectious disease (numbers needed to harm).

Despite smaller numbers, methotrexate initiation was found to be associated with a lower risk of infection than azathioprine in all sensitivity analyses that were designed to test different follow-up times (three and nine months), the exclusion of those with a history of dispensation of either methotrexate or azathioprine at any time point prior to a trial emulation, and various definitions of infectious disease. Similar estimates were also obtained after replication of analyses with modified Poisson regression albeit confidence intervals were wider than those estimated using targeted maximum likelihood estimation combined with data adaptive techniques. Unmeasured confounding was unlikely to explain the findings of the main analysis as indicated by the large E-value of 2.9. An association of a magnitude of 2.9 on the risk ratio scale would be needed between an unmeasured confounder with both treatment initiation and infectious disease within six months to explain away the observed favorable effect of methotrexate initiation compared to azathioprine initiation (risk ratio point estimate 0.57 from the main analysis).

5.2.6 Study VI: Risk and predictors of heart failure in sarcoidosis

In this study, 8574 individuals with sarcoidosis and 84 192 matched general population comparators were followed for heart failure for a median of 4.8 years (interquartile range 2.3, 7.6). As depicted in Table 16 (page 58), 204 individuals with sarcoidosis and 721 comparators without sarcoidosis were diagnosed with heart failure during follow-up. The incidence rate was higher in sarcoidosis compared to the general population (2.2 versus 0.7 per 1000 person-years). A 2.4-fold higher relative risk of heart failure associated with sarcoidosis was observed after adjusting for demographics and comorbidity (hazard ratio 2.43 [95% CI 2.06, 2.86]).

Hazard ratios of heart failure did not differ among strata defined by age at start of follow-up, sex, or sarcoidosis treatment status around diagnosis (*P*-value from a likelihood ratio test >0.05 for all comparisons). A higher hazard ratio was identified during the first two years of follow-up compared to the rest (hazard ratio 3.74 versus 1.86; *P* <0.001) and in individuals without compared to those with a history of ischemic heart disease or acute myocardial infarction before sarcoidosis diagnosis or the corresponding period for matched comparators (hazard ratio 2.74 versus 1.70; Table 16, page 58).

Table 16 | Risk of heart failure in sarcoidosis compared to the general population.

Adjusted hazard ratios of heart failure comparing sarcoidosis to the general population, overall and stratified by age at start of follow-up, sex, sarcoidosis treatment status around diagnosis, time since start of follow-up, and history of ischemic heart disease or acute myocardial infarction.

	Sarcoidosis	General population	Hazard ratio^a (95% CI)
Overall sarcoidosis	204/8574 (2.2)	721/84 192 (0.7)	2.43 (2.06, 2.86)
Age at start of follow-up, years			
18–44	18/3699 (0.8)	53/36 549 (0.2)	2.79 (1.63, 4.78)
45–64	68/3394 (3.3)	225/33 578 (1.1)	2.53 (1.92, 3.34)
65–85	118/1481 (19.0)	443/14 065 (6.6)	2.32 (1.88, 2.87)
Sex			
Female	90/3824 (1.8)	329/37 606 (0.6)	2.25 (1.77, 2.86)
Male	114/4750 (2.4)	392/46 586 (0.8)	2.59 (2.08, 3.21)
Sarcoidosis treatment around diagnosis ^b			
Treated	62/2696 (3.0)	161/26 479 (0.7)	3.35 (2.48, 4.53)
Untreated	79/3889 (2.1)	234/38 120 (0.6)	2.71 (2.08, 3.52)
Years since start of follow-up			
≤2	98/8574 (1.2)	223/84 192 (0.5)	3.74 (2.93, 4.77)
>2	106/6585 (1.0)	498/65 949 (0.2)	1.86 (1.50, 2.31)
History of ischemic heart disease or acute myocardial infarction			
Yes	37/339 (26.9)	186/2735 (13.6)	1.70 (1.19, 2.44)
No	167/8235 (2.1)	535/81 457 (0.6)	2.74 (2.28, 3.29)

CI = confidence interval.

Data are number of heart failure diagnoses after start of follow-up/number of individuals at risk (incidence rate per 1000 person-years estimated using Poisson models adjusted for age, sex, and region of residence) unless otherwise stated.

^aHazard ratios were estimated using Cox proportional hazards models with attained age as the time scale adjusted for age at start of follow-up, sex, and region of residence, country of birth, education, civil status, calendar period, healthcare visits within two years before inclusion, history of hypertension, diabetes mellitus, dyslipidemia, heart valve disease, atrial fibrillation, chronic obstructive pulmonary disease, chronic kidney disease, alcohol-related disorders, and autoimmune disease.

^bEvaluated in individuals with sarcoidosis and their matched general population comparators who entered the cohort in 2006 and onwards for whom prescription dispensation data was available in the Prescribed Drug Register (established in July 2005).

A hazard ratio of 2.30, similar to that of the main analysis, was found when heart failure diagnoses during the first six months of follow-up were disregarded. A small attenuation of the association between sarcoidosis and heart failure was observed when secondary discharge diagnoses were considered in the heart failure definition (hazard ratio 1.87).

Table 17 | Predictors of heart failure in sarcoidosis.

Mutually adjusted hazard ratios of heart failure for demographic and clinical predictors evaluated at start of follow-up.

Predictor	Heart failure (n=204)	No heart failure (n=8370)	Hazard ratio^a (95% CI)
Age at sarcoidosis diagnosis ^b , years	65 (12.9)	49 (14.4)	1.9 (1.7, 2.2)
Male	114 (56)	4636 (55)	1.5 (1.1, 2.0)
Education ^c , years			
≤9	98 (48)	1781 (21)	1.9 (1.3, 2.9)
10–12	73 (36)	4129 (49)	1.3 (0.9, 2.0)
≥13	33 (16)	2460 (29)	1.0 [Referent]
Hypertension	107 (52)	1789 (21)	1.2 (0.9, 1.7)
Diabetes mellitus	58 (28)	609 (7)	2.5 (1.8, 3.6)
Chronic kidney disease	20 (10)	173 (2)	1.4 (0.8, 2.2)
Alcohol-related disorder	10 (5)	186 (2)	2.0 (1.0, 3.9)
Chronic obstructive pulmonary disease	11 (5)	225 (3)	1.1 (0.6, 2.1)
Heart valve disease	16 (8)	73 (1)	1.8 (1.0, 3.1)
Ischemic heart disease or acute myocardial infarction	47 (23)	343 (4)	1.4 (1.0, 2.1)
Atrial fibrillation	49 (24)	214 (3)	2.6 (1.8, 3.9)
Other arrhythmia or heart block	39 (19)	277 (3)	2.5 (1.6, 3.7)
Pulmonary hypertension	≤5 (2)	7 (<1)	Not estimable
Autoimmune disease	27 (13)	674 (8)	1.1 (0.7, 1.7)

CI = confidence interval.

Data are n (%) or mean (standard deviation) unless otherwise stated.

^aHazard ratios were estimated from a Cox proportional hazards model with years since sarcoidosis diagnosis as the time scale and covariates age, sex, region of residence, country of birth, education and all comorbidities in the table except for pulmonary hypertension due to small numbers.

^bEffect per 10-year increase.

^cCategory “≤9 years” includes <1% missing.

The results of the analysis of predictors of heart failure in sarcoidosis are shown in Table 17 above and Figure 14 (page 60). Individuals with sarcoidosis who were diagnosed with heart failure during follow-up were more likely to be older, of male sex, and have less years of completed education. Comorbidity was also more prevalent at start of follow-up in those diagnosed with heart failure. In a mutually adjusted Cox model, diabetes mellitus, atrial fibrillation and other arrhythmias (including ventricular arrhythmias, heart blocks, and/or non-fatal cardiac arrest) were found to be the strongest clinical predictors of heart failure diagnosis. Each of the three predictors was independently associated with an approximately 2.5-fold increased relative risk of heart failure. Considering their prevalence at start of follow-up (sarcoidosis diagnosis), they were estimated to account for about 20%, 16%, and 12% of heart failure diagnoses within two years from sarcoidosis diagnosis, respectively (attributable fraction; Figure 14, page 60).

Other clinical predictors (e.g., hypertension, heart valve disease, or chronic kidney disease) were weakly associated with a higher relative risk of heart failure in sarcoidosis. A composite variable including history of ischemic heart disease and acute myocardial infarction was associated with a 40% increased relative risk of heart failure (attributable fraction 7% at two years after sarcoidosis diagnosis).

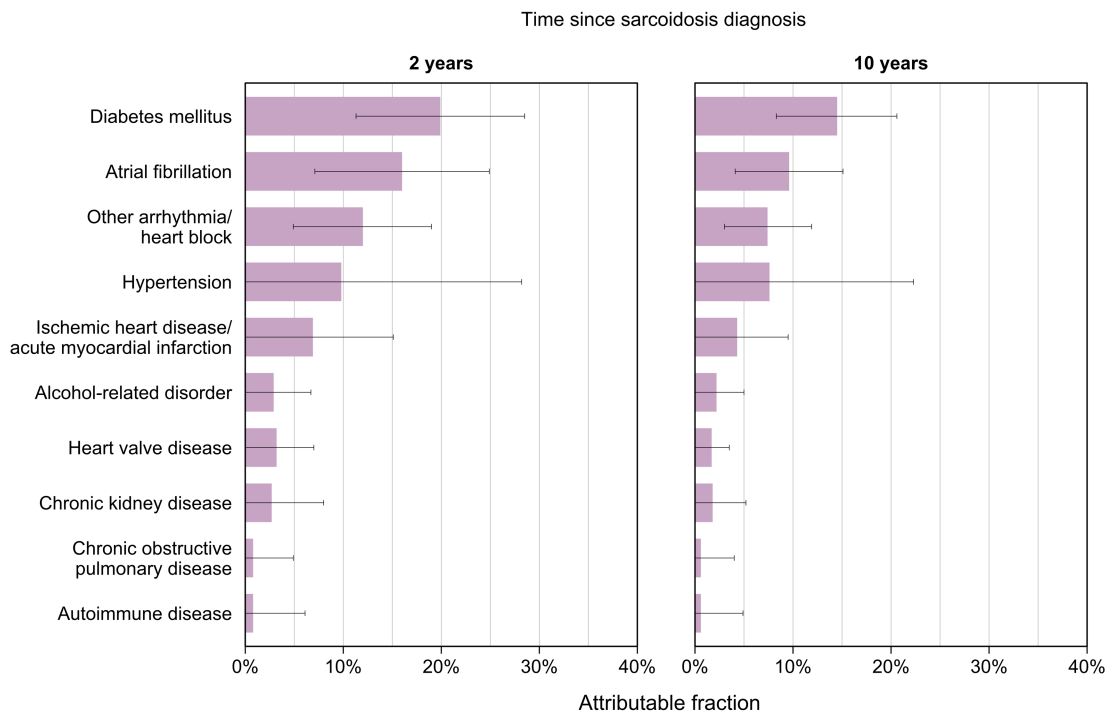


Figure 14 | Attributable fractions of heart failure diagnoses at two and 10 years after sarcoidosis diagnosis.

Evaluating clinical predictors also during follow-up resulted in slightly higher hazard ratios for heart failure. Treatment with immunosuppressants around sarcoidosis diagnosis did not appear to be associated with an increased risk of heart failure in analyses restricted to individuals with sarcoidosis diagnosed between 2006 and 2013 for whom prescription data in the PDR was available (hazard ratio 1.2 [95% CI 0.8, 1.7]). Similarly, the defined daily dose of systemic corticosteroids dispensed within six months before start of follow-up was not associated with heart failure during follow-up in treated individuals. However, a twofold higher relative risk of heart failure was found in those who received more than 300 defined daily doses of systemic corticosteroids within the first six months after start of follow-up compared to those treated with 150 defined daily doses or less during the same period (hazard ratio 2.1 [95% CI 1.1, 3.8]).

6 DISCUSSION

6.1 THE FINDINGS OF THIS THESIS IN CONTEXT

Six individual studies are included in this thesis which examined various aspects of sarcoidosis epidemiology including disease etiology and longer-term consequences of the disease on patients. This thesis and the individual studies serve as proof of the usefulness of register-based research in answering etiological and clinical questions in a relatively rare and heterogenous disease in which large numbers and robust methods are of utmost importance. Nonetheless, the data and methods used to conduct the individual studies in this thesis are by no means impervious to criticism. Potential issues and disadvantages of the data and methods used herein are extensively discussed next in this section (subchapter 6.2, page 69). First, this thesis' findings are interpreted and put in context of the available literature on each respective topic. The implications of this thesis' results for treating physicians, patients, and sarcoidosis researchers are discussed in Chapter 8 ("Points of perspective", page 77).

6.1.1 Risk factors for sarcoidosis

In Studies I and II, we examined whether familial disease and history of infectious diseases predispose to sarcoidosis. Study I was the largest to investigate the familial aggregation and heritability of sarcoidosis using a case-control-family design aiming to minimize misclassification bias originating from differential identification of relatives and their disease status dependent on whether a proband was a sarcoidosis case or a control.

Familial disease is a strong risk factor for sarcoidosis occurrence, but the heritability of sarcoidosis is lower than previously thought.

In Study I, we found that sarcoidosis clusters in families. The prevalence of familial sarcoidosis was estimated to be approximately 4%, somewhat lower than what was indicated in some hospital-based cohorts (pooled prevalence 9%) [231]. In relative terms, individuals with at least one first degree relative diagnosed with sarcoidosis in our cohort had a 3.7-fold increased risk of receiving the diagnosis of the disease compared to those with no relatives with sarcoidosis. The relative risk of sarcoidosis varied by the number of affected relatives and by kinship in a dose-response manner. Having two or more first degree relatives with sarcoidosis further increased the risk of sarcoidosis, whereas the familial relative risk was half of that (1.5) in those who had a half sibling with the disease. Familial relative risk being determined by the amount of 'familial exposure' to disease was an indicator of the implication in disease etiology of genetics, shared environmental factors, or both. Aside from some high risks of sarcoidosis attributed to exposure to some microbial agents [82,133,232], which are further discussed

below, familial disease is one of the strongest risk factors for sarcoidosis identified in the literature.

Although Study I is the largest study on familial aggregation to date, ACCESS was the first major attempt to examine the familial clustering of sarcoidosis. The few other studies on this topic [106,109,110] were either low-powered or did not include a control group, resulting in potentially unreliable estimates. Our main finding of a more than threefold higher relative risk of sarcoidosis associated with disease in first degree relatives was similar to that of the ACCESS [111]. However, there were several differences between our and findings from the ACCESS. In the ACCESS, the relative risk associated with having second degree relatives with the disease was higher than that for first degree relatives [111]. This observation does not conform to the notion that genetics might play some role in sarcoidosis etiology. One would expect to find higher relative risks with increasing genetic exposure in a proband-relative cluster. First degree relatives and probands share 50% of their genetic material while second degree relatives and probands share half of that (25%). In addition, an 18-fold increased relative risk of sarcoidosis was found in white Americans (2.8 in black Americans) [111], whom we perceived to be ethnically more similar to the Swedish-born probands and their relatives included in our study (familial relative risk 3.7). The reasons that could explain the differences among our study's findings and those of the ACCESS are challenging to identify. It is likely that small numbers in stratified analyses, potential differential misclassification due to sarcoidosis status in relatives reported by the proband, and the modeling approach for their case-control dataset that was based on non-conditional logistic regression in which the dependent variable was the relative's disease status, may have all contributed to the discrepancies outlined above.

Moreover and in contrast to the ACCESS [111], there was a tendency in our study for a higher familial relative risk of sarcoidosis in probands diagnosed with sarcoidosis at a younger compared to an older age, with a familial relative risk of 4.0 in those aged younger than 50 years at sarcoidosis diagnosis compared to 3.5 in probands 50 years or older (i.e., the mean age at sarcoidosis diagnosis in our cohort). A similar observation of a higher familial relative risk was made for probands with Löfgren's compared to non-Löfgren's disease when we restricted to individuals registered in the Karolinska Clinical Cohort for whom phenotypic data was available. These findings might suggest that diagnosis at younger age is an indication of stronger genetic susceptibility to disease (herein estimated by the familial relative risk) and Löfgren's being genetically distinctive and potentially more hereditary than non-Löfgren's disease [116]. However, unless these observations are replicated with larger numbers, they should be interpreted with caution because the varying sex distribution across age groups with males diagnosed at a younger age than females and the misclassification of the non-Löfgren's disease group, which was an amalgamate of several different sarcoidosis phenotypes, possibly contributed to those observations.

The threefold increased familial relative risk of sarcoidosis that we showed in the first part of Study I was an indication of the implication of shared risk factors for the disease within a

family. This finding motivated our analysis that aimed to quantify the potential additive genetic, shared and non-shared environmental components of the susceptibility to sarcoidosis. Therefore, in subsequent quantitative genetic modeling, we found that additive genetic factors could account for up to 39% of the susceptibility to sarcoidosis while the rest of the susceptibility to sarcoidosis was attributable to non-shared environmental factors.

Our estimate of sarcoidosis heritability is lower than what was previously indicated in two studies: one from the United States published in 1976 and a more recent study in twins from Finland and Denmark (heritability 60–70% and 66%, respectively) [106,113]. Assuming a somewhat similar prevalence of sarcoidosis and distribution of environmental exposures among Sweden, Finland, and Denmark, we would expect a somewhat similar heritability between these Scandinavian countries. Similarly, although the study from the United States was based on a sample of black Americans in whom the prevalence of sarcoidosis is among the highest in the world, previous studies have consistently shown that familial relative risks of sarcoidosis in this population average around 3.0 [110,111]. Considering these factors, one would expect a lower heritability than that of 60–70% that was found in the black American population [106,206]. Other factors related to the different designs and methods used in these two studies compared to ours and the small number of probands and exposed first degree relatives in the first [106], and twin pairs concordant for sarcoidosis in the second [113], may have led to some overestimation of heritability in those studies [205].

Preclinical sarcoidosis increases the risk of infection before diagnosis rather than the opposite.

In Study II, we did not find enough evidence in support of the notion that clinical infectious diseases are etiologically linked with sarcoidosis. Specifically, in Study II, we examined the relative risk of sarcoidosis associated with a history of infectious disease diagnosed during a hospital or an outpatient visit (or through antimicrobial medication dispensations in a subsequent analysis). We found that history of infectious disease evaluated at least three years before sarcoidosis diagnosis was associated with a small increase in the relative risk of sarcoidosis in the future (odds ratio 1.2). A similar relative risk was observed when different definitions of infectious disease aiming to capture varying severity of infections were used. Common infections, such as those of the upper respiratory and urinary tracts were likely to be behind the small increased relative risk of sarcoidosis. Testing latency periods between exposure ascertainment and sarcoidosis diagnosis (first visit in the NPR) ranging from one to seven years before sarcoidosis diagnosis resulted in a similar relative risk of sarcoidosis (approximately 1.2). A 50% higher risk of sarcoidosis was found, however, when history of infectious disease was ascertained right before sarcoidosis diagnosis (odds ratio 1.5).

Study II remains the first epidemiological study to have comprehensively examined the relative risk of sarcoidosis associated with history of infectious disease. This does not allow for adequate triangulation of the evidence presented here. A recent register-based study from Taiwan indicated that the risk of sarcoidosis is eightfold increased in patients with compared to those without a prior diagnosis of tuberculosis [82]. We could not replicate this finding in our sample.

As discussed above (see subsection 2.3.2.1; page 16), infectious agents were and are still regarded to be the most prominent candidates to explain the onset of sarcoidosis in genetically susceptible individuals [132,144]. Among those is infection by *Mycobacterium tuberculosis* and *Cutibacterium acnes* for which meta-analyses of molecular studies found relative risks of sarcoidosis as large as 16 and 20, respectively [232,233]. For this reason, one would expect to observe much high relative risks of sarcoidosis associated with infectious disease, a pattern of increasing relative risks the closer a patient was approaching sarcoidosis diagnosis, and potentially higher relative risks associated with infectious diseases at sites related to sarcoidosis pathophysiology such as the lower respiratory system and skin. Against expectations, none of these observations were made in this study. It should be emphasized, however, that our data could merely capture clinical manifestations of an infection by those and other agents. Latent or indirect triggering of sarcoid granulomatous inflammation by these agents (e.g., by molecular mimicry), which was suggested in some studies [132,144], could not be examined in Study II.

Our observations from Study II highlighted a rather common but often disregarded problem with etiologic studies, that of reverse causality (differential misclassification). From the outset of this study, we are aware of the fact that long preclinical disease, which characterizes several inflammatory diseases and possibly sarcoidosis [58,234–237], could lead to small and spurious associations between infectious disease history and sarcoidosis. To estimate how much of reverse causation bias (differential misclassification of the exposure) would be required to explain the observed weak associations, we conducted probabilistic bias analyses (simulations). We showed that if infectious disease in about one out of 10 future sarcoidosis cases was due to preclinical sarcoidosis it was large enough to explain the odds ratio of 1.2 of sarcoidosis associated with history of infectious disease. As we have previously shown, about 10% of individuals may present signs of preclinical disease years before sarcoidosis diagnosis, which could give rise to infections during that period, including increased utilization of healthcare services, dispensation of medications, receipt of sarcoidosis-related diagnoses such as uveitis, and sick leave absence [58,65].

Several questions were raised by the findings of Study II. Does the presence of reverse causation bias entail that infectious agents are not etiologically linked with sarcoidosis? Does this bias only affect register-based epidemiological investigations or extends to molecular studies? What is a possible mechanism for this bias? It is challenging to provide concrete answers to those questions as evidence is lacking. The mere presence of reverse causation bias in current molecular or epidemiologic studies does not preclude the possibility that an

infectious agent can trigger granulomatous inflammation either directly or indirectly by triggering an autoimmune-like response in a susceptible individual [144,238]. We know from Study I (on familial aggregation of sarcoidosis) and molecular genetic studies [116] that individuals who develop sarcoidosis have a genetic predisposition to the disease. One may speculate that the processes that underlie sarcoid inflammation develop gradually and diagnosis is triggered by either an abrupt endogenous or exogenous event, or when a certain threshold is reached. A similar mechanism has been proposed in other inflammatory diseases. In rheumatoid arthritis and systemic lupus erythematosus, production of auto-antibodies predates disease diagnosis by almost a decade in some cases [234,236]. It remains to be seen whether there is any evidence in favor of this potential mechanism. If true, however, that would entail that findings from both epidemiologic and molecular etiologic studies should be interpreted with great caution and future studies should find neat ways to account for potential bias due to reverse causation.

6.1.2 Long-term consequences of sarcoidosis

In Studies III to VI, we examined the burden of important and debilitating long-term outcomes in patients with sarcoidosis, namely mortality, infection, and heart failure.

The risk of early death is higher in sarcoidosis compared to the general population, especially for severe sarcoidosis.

In Study III, we showed that individuals with sarcoidosis have a 1.6-fold increased risk of all-cause death compared to the general population. Higher relative risks for mortality in the range of 1.6 to 2.4 for sarcoidosis overall were shown in all [97,180] but one of cohort studies [78] that either predated or followed this study. The exception was the study that used data from the Olmsted County cohort from the United States, which showed no association between sarcoidosis and all-cause mortality [78]. As shown in Figure 15 (page 66), pooling estimates from all available cohort studies on sarcoidosis mortality using a random-effects model resulted in a relative risk of 1.57, similar to the one shown in our study for overall sarcoidosis.

This is the first population-based study with enough power to examine risk stratification by demographic or sarcoidosis-specific factors. Relative risks of mortality did not markedly differ across age groups or between males and females. These findings are in line with two hospital-based studies on predictors of mortality in individuals with sarcoidosis from the United States [239] and France [240]. Although a higher female to male ratio in terms of mortality was found in cross-sectional studies based on death certificates from both countries

[176,241], this observation could not be replicated in either of the studies on predictors [239,240] or in our cohort.

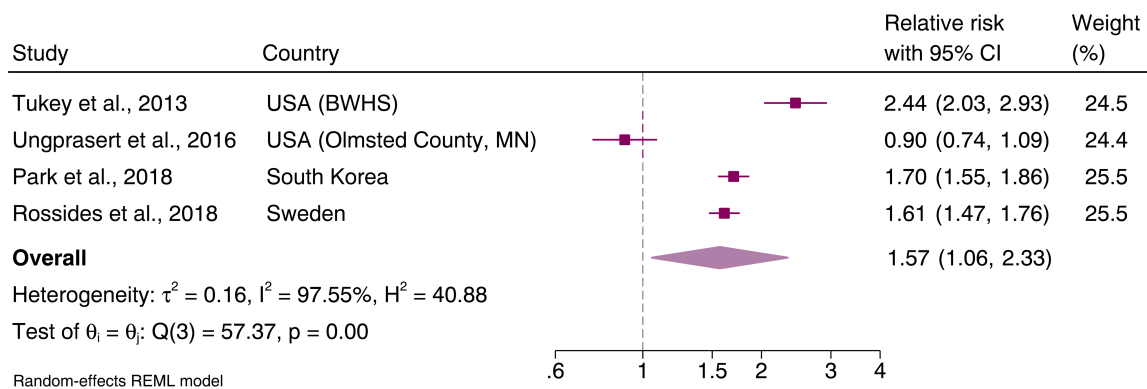


Figure 15 | Relative risk of all-cause death associated with sarcoidosis compared to the general population.
(BWHS = Black Women's Health Study; MN = Minnesota; CI = confidence interval)

To investigate whether relative risks of all-cause death varied by sarcoidosis severity, we stratified our patient population by treatment status around the time of diagnosis as immunosuppressant treatment is recommended in patients with severe signs or symptoms, or those with progressive (and likely chronic) disease [2]. We observed a more than twofold increased relative risk of mortality in those who received treatment around diagnosis while the association between sarcoidosis and all-cause death was almost null in those who did not receive treatment during the same period.

This finding is in line with observations in the two aforementioned studies on mortality predictors in which various proxies of sarcoidosis severity based on lung imaging (e.g., stage IV disease on chest X-rays, signs of fibrosis in high-resolution computed tomography scans, etc.) were associated with higher relative risks of mortality [239,240]. The majority of those patients were presumably under treatment with immunosuppressants, most commonly corticosteroids [240] (treatment status was not reported in [239]). Interestingly, the risk of mortality in black American patients was more than twofold higher than that in white Americans [239], an observation that may explain the higher relative risk of all-cause death observed in the investigation of mortality using data from the Black Women's Health Study [180] compared to what we showed in our cohort from Sweden (relative risk for sarcoidosis overall 2.4 versus 1.6; Figure 15).

A long list of comorbidities and sarcoidosis manifestations are believed to contribute to these heightened risks or early death. We showed in this thesis that patients with sarcoidosis are at higher risks of serious infection, sometimes recurrent, and heart failure compared to the general population. These are two complications that are associated with premature death in sarcoidosis

and other diseases [192,242]. Indeed, in Study III, we found an excess of cases of heart failure and infectious diseases in deceased individuals with sarcoidosis than those without. Moreover, sarcoidosis-related manifestations and long-term complications including cardiac disease [21,95,243], sarcoidosis-associated pulmonary hypertension [244,245], pulmonary fibrosis [246], and others might increase risks for early death. Last, further investigation of the role of morbidity associated with treatment (e.g., diabetes and/or osteoporosis in those treated with oral corticosteroids) is required to enable evidence-based decisions on who, when, and how to treat.

Study IV showed that patients with sarcoidosis have an overall 81% higher risk of being hospitalized with a serious infection compared to individuals from the general population without sarcoidosis. Relative risks were markedly higher during the first two years since diagnosis and tapered off later during the follow-up. Recurrent infections were also more common in individuals with sarcoidosis than in the general population. In contrast to Study III on sarcoidosis mortality, we found that relative risks of serious infection were notably increased irrespective of treatment status around the time of sarcoidosis diagnosis, a marker of sarcoidosis severity. There were differences between the two groups determined by sarcoidosis treatment status, with risks in treated patients being threefold higher than in the general population compared to 1.5-fold in untreated patients with sarcoidosis.

The relative risk of infection is high in sarcoidosis, especially within the first two years from diagnosis. Second line treatment methotrexate is safer than azathioprine in terms of infection.

A similarly higher overall relative risk of serious infection associated with sarcoidosis was found in the cohort from Olmsted County, Minnesota in the United States [185]. These findings are not surprising given the fact that a high burden of serious infection was observed in several other inflammatory or autoimmune diseases [247–250]. Higher relative risks during the first two years since diagnosis conform to what we know about the clinical course of sarcoidosis. During the first couple of years after diagnosis, most patients have active disease, about 40% receive immunosuppressive treatment, some undergo invasive examinations, and are, in general, in frequent contact with the healthcare system, which increases somewhat the likelihood of them receiving any diagnosis.

Systemic corticosteroid use has been linked with both short- and long-term adverse outcomes, among others, infections [251–254]. Higher relative risks of serious infection in patients with sarcoidosis who received those medications indicate a possible contributory role of treatment to these risks. In the absence of a severity index other than immunosuppressive therapy, however, we could not disentangle risks of infection due to corticosteroid treatment or

attributable to sarcoid inflammation for which such treatment is an indication. Nevertheless, in contrast to some other studies [255], the absence of any signals of higher occurrence of opportunistic infections in these patients showed that immunosuppression was not severe.

In terms of second line treatments, our findings from Study V showed that methotrexate is associated with a 43% lower risk of infection at six months after treatment initiation compared to azathioprine, an alternative treatment. Despite some variation, the advantage of methotrexate over azathioprine in terms of infection was consistently present in several analyses that were designed to stress-test the association. Moreover, in Study V, we could adjust for sarcoidosis severity using corticosteroid treatment as a proxy as well as control for several other factors to minimize the possibility that treatment comparisons were in any way influenced by confounding by indication.

The comparative safety of methotrexate and azathioprine in sarcoidosis has not been examined in a randomized clinical trial and evidence from trials in other diseases was inconclusive because those were not designed to investigate safety outcomes [224,225,256–258]. Our findings, however, are in line with a descriptive study from two European centers that showed a lower proportion of infection in patients with sarcoidosis treated with methotrexate compared to azathioprine [61] and an observational study in patients with rheumatoid arthritis in which relative risks of serious infection were lower in methotrexate- than azathioprine-treated patients [259]. In addition, a recent randomized, placebo-controlled clinical trial showed that use of methotrexate was not associated with a higher risk of infection [260], which further suggests that a possible advantage of methotrexate compared to azathioprine may also be present in sarcoidosis.

**The relative risk of heart failure is notably high in sarcoidosis.
Is cardiac sarcoidosis to blame?**

In Study VI, we examined relative risks of heart failure associated with sarcoidosis and identified predictors of a heart failure diagnosis in sarcoidosis. We showed that the risk of heart failure was 2.4-times higher in sarcoidosis compared to the general population and was even higher (relative risk 3.7) during the first two years after sarcoidosis diagnosis. The relative risk of heart failure associated with sarcoidosis is higher in individuals without compared to those with a history of ischemic heart disease (unstable angina or acute myocardial infarction; relative risk 2.7 versus 1.7, respectively).

These findings add to the body of literature on heart failure risks in sarcoidosis. Higher relative risks of heart failure in sarcoidosis were indicated in two recent studies [190,192]. Among those is a large Danish register-based study, which showed that heart failure risks are not only high in sarcoidosis compared to non-sarcoidosis controls particularly within the first

year from sarcoidosis diagnosis, they also lead to excess mortality in this group of patients [192]. The latter finding is in line with our observation in Study III on sarcoidosis mortality in which an excess of deaths associated with heart failure (as coded on death certificates) was observed in sarcoidosis compared to the general population.

Although our results for sarcoidosis overall were not entirely unexpected given this prior knowledge, our study highlighted several important characteristics of heart failure and its etiology in sarcoidosis that were not previously investigated. In absolute numbers, heart failure is more common in older individuals. However, in sarcoidosis, females and males of all ages are at higher risk of being diagnosed with heart failure. Unlike Studies III and IV, which showed markedly higher relative risks of mortality and infectious disease in individuals with presumably severe or progressive sarcoidosis (as approximated by treatment status around diagnosis), heart failure appears to affect all individuals with sarcoidosis irrespective of disease severity.

In addition, most heart failure cases in sarcoidosis as in the general population are diagnosed on a basis of ischemic heart disease. We identified, however, a higher relative risk of heart failure in individuals without ischemic heart disease, which suggests an interaction between the pathophysiologic mechanisms of sarcoidosis and heart failure. Indeed, our analysis of clinical predictors of heart failure indicated that a large proportion of heart failure cases in sarcoidosis could be attributed to cardiac arrhythmias (especially ventricular arrhythmias and severe heart blocks), which was interpreted as an indication of the critical role of cardiac sarcoidosis in these patients. This is hardly surprising, as heart failure is considered to be the end stage manifestation of unmitigated and non-fatal cardiac involvement in sarcoidosis. Presence of heart failure increases the probability of diagnosis of cardiac sarcoidosis in a patient with sarcoidosis [22].

6.1.3 Transportability of findings

All individual studies in this thesis were based on a large and representative cohort of individuals with sarcoidosis and included several stratifications on potential effect modifiers (demographic and sarcoidosis-related factors) to increase the generalizability and applicability of our findings. The results of individual studies are generally transportable to other populations that are similar to this cohort from Sweden in terms of ethnic background, access to quality healthcare, sarcoidosis severity, and treatment status.

6.2 METHODOLOGICAL CONSIDERATIONS

6.2.1 Registers in clinical epidemiologic research: friend (and foe?)

In this thesis, a large linkage of Swedish population-based registers was used. This resource of secondary data provided enough power to conduct most analyses, minimized selection bias due

to non-random loss to follow-up, and greatly reduced time and costs associated with collection and management of primary clinical or demographic data. Despite the obvious advantage of these registers, a number of limitations need to be commented on. Clinical variables and results of examinations are not registered in these resources, which resulted in some misclassification of exposures, outcomes, and confounders in the studies in this thesis. For example, we did not have information on the means of diagnosis of sarcoidosis as results of biopsies, bronchoalveolar lavage, and of other diagnostic examinations were not available in the NPR. That motivated the use of bias analyses for non-differential sarcoidosis misclassification in Studies I to III, and the conduct of a validation study in a sample of patients with sarcoidosis that showed a high (94%) positive predictive value [219].

Moreover, the lack of detailed clinical information on the study population did not allow us to provide more clinical context for diseases that were examined either as exposures or outcomes. We could not group patients with sarcoidosis according to phenotype or organ involvement (e.g., pulmonary, skin, cardiac, syndromic disease, etc.) that would have been useful, especially in Study VI on heart failure in which knowledge of cardiac sarcoidosis diagnoses would have allowed us to distinguish between heart failure with reduced or preserved ejection fraction, a common clinical distinction that determines therapy. In some instances, we used phenotypic information (Löfgren's versus non-Löfgren's disease) from our local clinical cohort at Karolinska to complement the main analyses, but data was not always adequate in number to perform all intended analyses. Nevertheless, definitions of exposures and outcomes used in this thesis were of proven high validity [228].

High validity does not mean high sensitivity. The NPR is a nationwide source of medical diagnoses but does not include diagnoses from primary care. For sarcoidosis in which most patients are diagnosed and followed-up in secondary care that would result in small misclassification. We have likely missed patients with diagnoses such as diabetes, hypertension, infection, and heart failure for which some patients are cared for entirely in primary care, data from which we did not have. Lack of primary care data in our database would result in some differential misclassification of the outcome in Studies IV and VI under the assumption that individuals with sarcoidosis were more likely to be referred to secondary care for the diagnosis of either infectious disease or heart failure compared to comparators without sarcoidosis. To minimize that risk, we used, when possible, information on dispensation of prescribed medications obtained from the PDR. Primary care data would have provided some information on lifestyle factors (e.g., smoking and obesity) that were potential confounders of almost all of the associations examined in this thesis. Due to the lack of data on lifestyle factors, we resorted to simulation analyses to test the robustness of the examined associations in the presence of unmeasured confounding.

A major hinder to sarcoidosis research, and by extension, to the studies included in this thesis is the lack of markers or indices of sarcoidosis severity that are either sensitive, specific, or applicable to all manifestations of sarcoidosis. A few potentially relevant markers of severity either based on laboratory or imaging examinations (e.g., spirometry parameters or fibrosis on

a high-resolution computed tomography scan) or patient-reported symptoms (e.g., dyspnea and fatigue) that could inform our analyses could not be obtained from the NPR (or any other source). To approximate disease severity, we used treatment with immunosuppressants around the time of diagnosis and in some cases when numbers allowed, Löfgren's and non-Löfgren's disease status recorded for some patients diagnosed at Karolinska University Hospital. The prescription of immunosuppressive treatment likely correlated well with the inflammatory load in a patient and that hampered the separation of effects owing to treatment or sarcoid inflammation on patient outcomes such as mortality, infectious disease, and heart failure.

Considering these limitations, it may appear impossible to conduct epidemiologic research using large registers. The studies in this thesis prove the opposite. Registers and their wealth of longitudinal information serve a very useful purpose, especially in relatively rare and under-researched diseases like sarcoidosis for which collection of primary data on a large scale is either infeasible or uneconomical. When complemented with demographic and clinical data, registers form a powerful resource that can be used to answer a wide variety of research questions related to disease etiology and patient outcomes.

6.2.2 Limitations of concepts and methods

Several different methods were used in the individual studies, each with advantages and disadvantages. The following paragraphs summarize some of the most critical limitations of select statistical and epidemiological concepts and methods used in this thesis.

In Study I, we used **biometric modeling** to estimate the proportion of variance in sarcoidosis risk in the Swedish population that was attributable to additive genetic effects (heritability), and shared and non-shared environmental factors. Quantitative genetic analyses require large numbers, which was reflected in the wide confidence intervals of our heritability estimate. Furthermore, this analysis was based on several assumptions that are worth mentioning. The liability-threshold model, the theoretical ground of such modeling, assumed a normal distribution for the susceptibility to sarcoidosis in the population that gave rise to the observed cases, that is, individuals who exceeded an unknown threshold [112]. This assumption is largely untestable. In addition, the model assumed mating in the population occurred at random, genetic effects other than additive were not present, and the prevalence of sarcoidosis was accurately measured. The latter is expected to hold, while violations to the first two strict assumptions have already been described [130], but their impact on our findings remains unknown and difficult to predict.

Heritability is an estimate of the variance in liability, as previously emphasized, an underlying, unobserved theoretical scale that can be attributed to additive genetic effects. It is a complex and not a readily intuitive measure to interpret. It is useful for academic discussions like the one in this thesis, but has little relevance when estimating the impact of interventions or communicating risks to patients [261]. To elucidate this point, consider the example from this thesis: an estimate of sarcoidosis heritability of 39% should not be interpreted as a 39%

reduction of the risk of sarcoidosis in the population if all genetic factors associated with sarcoidosis were to be effectively eliminated (in an obviously imaginative scenario). A measure of the latter is the population attributable fraction [261,262], which is discussed further below.

In Study II, we used a **high-dimensional propensity score** to capture as many confounding factors as possible that would have otherwise been left unmeasured. Our concern in Study II was that such unmeasured factors would inflate estimates of the association between infectious disease and sarcoidosis due to potential reverse causation bias [207]. The scope of a high-dimensional propensity score is to make efficient use of the multiple layers of healthcare data [221]. Propensity scores of this kind originated in pharmacoepidemiologic research [221] and are still commonly used in that field. There are a few issues with high-dimensional propensity scores that are worth mentioning. Most pharmacoepidemiologic studies are prospective and propensity scores can be readily estimated because data on exposure is collected independent of the outcome and represents the population of interest. That was not the case in our case-control study in which the exposure distribution did not reflect a random sample from the population due to the outcome-heavy sampling. We therefore opted to model the propensity for exposure (history of infectious disease) in the control sample [263] assuming no large bias was introduced due to matching because sarcoidosis is relatively rare in the general population [76].

Furthermore, it is unknown how a high-dimensional propensity score should be best handled in the analyses. Options include stratification by the score, matching on the score, use of the score in inverse-probability weighting, or merely adjust for it in traditional regression analyses, commonly in deciles [221]. We chose the latter, which was the simplest and probably most feasible and appropriate for our analysis. Another issue with high-dimensional propensity scores is the multiple testing involved in the creation of the score. Although the high-dimensional propensity score was shown to be useful for confounding adjustment in some situations [264], further stress tests are needed to prove its effectiveness and statistical properties, particularly in comparison to more modern data adaptive techniques.

In study V, we used a **target trial emulation** design to study the comparative safety of methotrexate and azathioprine in terms of infection. Target trial emulation with observational secondary data is a relatively new take on causal inference studies [204]. The core premise of this method is that careful planning of a causal inference study by outlining the ideal clinical trial that would answer the well-defined research question of interest will prevent the researcher from committing common errors in causal inference, namely mismatch between question and study design or analysis, selection bias, and/or immortal time bias [202,203]. Although the benefits of target trial emulation are obvious in contrast to traditional pharmacoepidemiologic or other causal inference studies, the method is not panacea and cannot replace all inference originating from properly conducted randomized controlled trials. Potential issues such as differential misclassification of the outcome due to the inability to blind participants and assessors to the administered treatment and confounding by indication, cannot always be prevented or corrected for at the analysis stage. It has, however, the potential to contribute with

timely and more generalizable evidence [202] when the target trial cannot be performed due to reasons related to ethics or lack of resources.

Also in Study V, we used a **doubly robust estimator** from the targeted maximum likelihood estimation framework [216] for the risk ratio that compared to occurrence of infection at six months after initiation of methotrexate compared to azathioprine. Estimation of the exposure, outcome, and missingness models was performed using an ensemble of models based on data adaptive techniques in addition to traditional regression algorithms. Except for more conservative inference, we did not observe any difference in our conclusions that were based on targeted maximum likelihood estimation over a traditional analysis conducted using modified Poisson models. Nonetheless, the targeted maximum likelihood estimation framework is a rapidly growing field of research and future developments will enable and test the usefulness and appropriateness of this framework in several more complex applications.

In Study II (infectious diseases as risk factors of sarcoidosis) and particularly in Study VI (heart failure risk and predictors), the concept of **population attributable fraction** [201] was used. This corresponds to Greenland and Robin's "excess fraction" [265]. Despite the fact that the attributable fraction is simple to understand, estimate, and communicate [266], it is often misinterpreted [267]. In Study VI, an attributable fraction for diabetes of 20% was estimated assuming all other risk factors were constant. Using diabetes as an example, the attributable fraction is an estimate of the proportion of heart failure cases (or average risk for heart failure), in this case 20%, that would be eliminated from the population if diabetes is completely eradicated by an intervention (over a defined time interval). The opposite is wrong, however. Twenty percent of heart failure cases will not have diabetes at diagnosis. Another common misconception is that a population attributable fraction should sum up to 100% when more than one risk factors are simultaneously investigated, like in Study VI. In studies in which exposures are not mutually exclusive, the population attributable fraction is not expected to sum up to 100% [261,267].

In several of the included studies, we conducted **probabilistic bias analyses** to quantitatively assess the impact of potential differential and non-differential misclassification and unmeasured confounding. Bias parameters (e.g., prevalence of smoking in exposed and unexposed to sarcoidosis) that were included in our simulations were informed by carefully examining external information given the lack of internal validation data. Quantitative bias analyses represent simple scenarios in which one or two biases are likely at play, potentially ignoring complex mechanisms or group-specific variation (e.g., smoking prevalence may vary by sex or calendar time). They are nevertheless preferable to qualitative discussion of biases because they allow for direct quantification of the effect of potential biases that is based on explicitly defined assumptions about the strength and direction of effects that are open to scrutiny by the reader [209].

7 CONCLUSIONS

Overall, this thesis and the included individual studies showed that register-based studies contributed to a better understanding of the epidemiology of sarcoidosis. Using a large linkage of health and administrative sources complemented with clinical data is an efficient way of answering questions pertaining to the etiology of a relatively rare disease like sarcoidosis and to clinical questions on patient outcomes after diagnosis. The field of sarcoidosis epidemiology has not been exhausted by these studies; further studies with contemporary and multifaceted data sources are needed to reach the end goal: prevent sarcoidosis or reduce its burden in those who suffer from it.

Specifically, based on the findings in individual studies in this thesis, we concluded that:

- Sarcoidosis clusters in families and individuals with relatives with sarcoidosis are at higher risk of developing the disease. A large proportion (39%) of the susceptibility to sarcoidosis appears to be attributable to additive genetic effects in the population; albeit non-shared environmental risk factors seem to also contribute to sarcoidosis etiology and merit an equal attention in future research studies.
- Infectious diseases do not appear to be strong risk factors for developing sarcoidosis. On the other hand, preclinical sarcoidosis may be associated with a higher risk of infectious diseases spanning years before diagnosis in some cases. Future molecular and epidemiologic studies should consider reverse causation as a potential explanation of findings when the true onset of disease is unknown. Better methods to diagnose sarcoidosis early are also warranted.
- Individuals with sarcoidosis are at higher risk of mortality compared to the general population. The relative risk is markedly higher in individuals who receive immunosuppressant treatment around the time of sarcoidosis diagnosis because of severe or likely progressive disease. More effective treatments and prevention or early diagnosis of comorbidity might help reduce high risks of early death in these patients.
- Patients with sarcoidosis are more likely to develop one or multiple serious infections compared to the general population, particularly during the first two years since sarcoidosis diagnosis and if they receive immunosuppressants due to perceived severe sarcoidosis. High alert and prevention measures during the first years after diagnosis, particularly in treated patients, may be considered to reduce those risks.
- Compared to azathioprine, methotrexate is associated with a notably lower risk of infectious disease within six months after initiation of either therapy. Unless contraindications prohibit its use, methotrexate could be considered as first choice when a second line treatment is necessary in patients with sarcoidosis.
- The relative risk of heart failure is markedly high in sarcoidosis, especially in the first two years after diagnosis and in individuals without a history of ischemic heart disease. Diabetes and arrhythmias (some presumably due to cardiac sarcoidosis) are independent predictors of heart failure in sarcoidosis. A relatively large proportion of heart failure cases could be prevented if these were to be appropriately controlled.

8 POINTS OF PERSPECTIVE

The findings in this thesis have some important implications for the diagnosis and care of individuals with sarcoidosis and for future research. They are discussed below and summarized study-by-study in Table 18 (page 80).

The findings of Study I are useful in medical practice and for future research. In clinical practice, inquiring about a patient's familial exposure to the disease may increase the confidence in diagnosing sarcoidosis. Future research is warranted to establish whether addition of family risk as a parameter in the diagnostic process may increase the specificity of future diagnostic guidelines. In addition, the results presented in this study could be readily used to inform patients about the 'heredity' of their disease and potential risks for their first and second degree relatives to be diagnosed with sarcoidosis. A heritability of sarcoidosis in the magnitude of 39% signifies that a large component of the variation in disease risk in the Swedish population owing to non-shared environmental factors is amenable to potential interventions aiming to prevent the disease [231]. Although recent research efforts have largely focused on the genetic component of susceptibility to sarcoidosis in the population, for which we admittedly know little about, research to identify the environmental factors that are responsible for the occurrence of sarcoidosis merits equal interest and allocation of research resources.

In Study II, we examined the role of infectious disease in sarcoidosis etiology, one of the most commonly implicated risk factors of sarcoidosis. We found little evidence that clinically identifiable infectious disease is a large contributor to sarcoidosis risk. In contrast, we uncovered a significant limitation of molecular and epidemiologic research dealing with disease etiology. When the disease onset is unknown or is expected to vary by phenotype (e.g., acute in Löfgren's syndrome versus insidious in non-Löfgren's disease), great caution is needed in the design, analysis, and interpretation of research findings. It is imperative that future studies should make use of biologic samples or longitudinal clinical data that is obtained before the disease is diagnosed. Biobank samples offer a great opportunity for such research endeavors. If sarcoidosis onset predates diagnosis for years in some cases, findings from the preclinical phase will help us understand the timing of events and map the pathophysiologic mechanisms involved. Until then, triangulation of findings and meticulously conducted bias analyses may help quantify the extent of reverse causation bias and set boundaries to study conclusions.

Further epidemiologic studies in sarcoidosis etiology are warranted. Studies on familial co-aggregation of sarcoidosis and other diseases, for example, autoimmune disorders, can help elucidate on the pathophysiologic mechanisms that drive sarcoid inflammation and its similarities or differences with other pathways. Moreover, studies in siblings might provide information on how environmental exposures (e.g., smoking, diet, obesity, occupation, etc.) add to the increased risk of genetically susceptible individuals. Expanding on studies of geospatial clustering of sarcoidosis by studying how this varies by climate patterns, soil

exposures, occupation, and access to healthcare services could provide some much-needed insights to sarcoidosis etiology.

Our studies on the consequences of sarcoidosis highlighted that the disease is associated with higher relative risks of serious infection and heart failure as well as early death in some individuals. Overall and stratified results from our large studies based on a representative patient sample can be readily utilized in clinical practice to inform patients who inquire about the risks associated with their new diagnosis. Our findings also emphasize the fact that greater awareness is required among treating physicians to diagnose and treat comorbidity early to prevent longer-term sequelae. The belief that sarcoidosis does not lead to major consequences in the majority of patients who are expected to recover within two to five years does not appear to be true or, at least, does not apply to all outcomes or all patients.

The majority of patients (about 60%) who do not have severe sarcoidosis at diagnosis and their disease is presumably less likely to progress to a chronic state may be reassured that sarcoidosis is unlikely to affect their life expectancy. It may affect, however, their quality of life as they might be at risk of debilitating outcomes, among others, multiple hospitalizations for serious infections and/or heart failure. Special attention is needed for a relatively large number of patients who present with severe disease and are less likely to completely recover within two to five years after their diagnosis. In these patients, risks of serious infection and heart failure are considerably higher, especially around the time of diagnosis and may result in higher risks of early death. In fact, the risk of all-cause death at five years since diagnosis in individuals with a severe disease phenotype is on average 7%, markedly higher than those with perceived uncomplicated or stable disease (five-year mortality risk 2%).

An array of diagnostic and preventive interventions could be useful in individuals with sarcoidosis at the highest risk of developing negative outcomes. Interventions should be applied early during the course of their disease and ideally target those who are more likely to benefit out of those. Examples include: (1) early detection of sarcoidosis manifestations (e.g., cardiac sarcoidosis), (2) induction of remission with safer and more effective pharmaceutical treatments than oral corticosteroids, (3) shift in preference towards safer second line treatment options, and (4) early detection of comorbidity such as diabetes, ischemic heart disease, heart failure, and infections, and prescription of primary and secondary preventive measures accordingly. Little evidence is currently available on the effectiveness of such measures and transportability of best practices from other inflammatory or autoimmune diseases may not always be warranted due to differences in age and sex distributions, pathophysiology, and severity among these diseases and sarcoidosis. Although some studies are already in development (e.g., a clinical trial designed to examine the effectiveness of methotrexate as first line treatment [268]), more research is needed to establish which of the aforementioned interventions are both feasible and effective.

Sarcoidosis is a very heterogenous and unpredictable disease in terms of prognosis, which largely impairs our efforts to target patients in need of preventive interventions. Healthcare and demographic data from large registers may be used to map trajectories of patients after

diagnosis and indicate factors that predict favorable or adverse outcomes already at diagnosis. Such data combined with phenotypic, clinical, genetic, and other molecular markers could notably improve our ability to predict a patient's likely clinical course at diagnosis and facilitate decisions on treatment initiation at the time of diagnosis or ideally throughout the course of disease.

In summary, this thesis provides evidence that can be used in daily clinical practice to inform treating physicians and patients. In addition, by filling research gaps in fundamental albeit less studied areas in this field, the studies in this thesis unlock new and exciting possibilities for future research. It is now apparent that the next study in the field of either sarcoidosis etiology or patient care should be large and collaborative, crossing boundaries both among countries and fields. Findings from translational research should meet evidence from clinical epidemiologic research and vice versa. Studies based on administrative databases have several advantages but may be limited if not complemented by systematic in parallel collection of clinical data. A combination of all worlds is needed to reach the end goal of sarcoidosis prevention and effective personalized care.

Table 18 | Key implications of the findings from the individual studies in this thesis.

Study	I am a treating physician...	I am a patient with sarcoidosis...	I am a sarcoidosis researcher...
I. Familial aggregation of sarcoidosis	<ul style="list-style-type: none"> About 4% of patients with sarcoidosis have a first degree relative with the disease Depending on kinship, patients are more likely to have sarcoidosis if they have a relative diagnosed with the disease Disease etiology is multifactorial; 39% of the susceptibility due to genetic variation in the population 	<ul style="list-style-type: none"> The chances of receiving the diagnosis are higher if I have a known relative with the disease and depends on our kinship Chances depend on relatedness (kinship), but not my or the sex of my relative Sarcoidosis is not determined completely by genetics (heredity); environmental factors likely play an important role 	<ul style="list-style-type: none"> Does familial disease determine sarcoidosis phenotype, disease severity, treatment response, or comorbidity? Does sarcoidosis and other diseases or traits (e.g. autoimmune disorders) tend to co-aggregate? Which are the genetic and environmental factors that predispose to sarcoidosis occurrence?
II. Infectious diseases as risk factors for sarcoidosis	<ul style="list-style-type: none"> There is no strong evidence from epidemiologic studies that a particular infectious disease predisposes to sarcoidosis History of infection before a patient is diagnosed is likely a marker of long preclinical disease 	<ul style="list-style-type: none"> Having a history of infectious disease diagnoses before sarcoidosis was likely due to slow-burning sarcoid inflammation that later became symptomatic and got diagnosed There is no clear evidence from epidemiologic studies showing that a particular infectious agent triggered my disease 	<ul style="list-style-type: none"> Attention to the presence of reverse causation bias is needed during design, analysis, and interpretation of findings of my molecular or epidemiologic study Sensitive and specific markers that could facilitate early diagnosis are needed
III. Mortality in sarcoidosis	<ul style="list-style-type: none"> In patients with mild, non-progressive disease, sarcoidosis is likely not to affect life expectancy In patients with severe or progressive disease who need treatment at diagnosis, mortality risks are 2.3-fold higher than those in the general population Mortality risks increase by age, but are similar in female and male patients 	<ul style="list-style-type: none"> Sarcoidosis does not affect my life expectancy if I have mild and uncomplicated disease that no immunosuppressant treatment is needed for For me who has more severe sarcoidosis and need such treatment, the risk of death is 7% at five years after diagnosis versus 2% in those without sarcoidosis or mild disease Risks for death increase as I age, but are similar in women and men 	<ul style="list-style-type: none"> Identify the factors that lead to higher risks of early death in some patients but not others Examine the effectiveness of interventions to control disease activity and comorbidity

Table 18 | Key implications of the findings from the individual studies in this thesis (continued).

Study	I am a treating physician....	I am a patient with sarcoidosis....	I am a sarcoidosis researcher...
IV. Risk of serious infection in sarcoidosis	<ul style="list-style-type: none"> • Patients are at high risk of serious infections, sometimes recurrent • The risk of serious infection is considerably higher within the first two years after diagnosis and in patients who receive immunosuppressants around the time of diagnosis 	<ul style="list-style-type: none"> • I am at higher risk of getting hospitalized with an infection than an individual without sarcoidosis and sometimes infections may be recurrent • More caution is needed if I receive immunosuppressants and during the first two years after sarcoidosis diagnosis. I should consult my physician to decide preventive strategies that are best suited for me 	<ul style="list-style-type: none"> • Strategies to lessen the burden of serious infection such as vaccinations, disease control with the lowest dose of immunosuppressant possible, change to a safer treatment, and prophylactic antibiotic use warrant further study
V. Infection in methotrexate versus azathioprine for sarcoidosis	<ul style="list-style-type: none"> • Methotrexate use in sarcoidosis is associated with a lower risk of infection than azathioprine • Methotrexate could be considered as first choice in patients who need a second line treatment for sarcoidosis unless contraindications exist 	<ul style="list-style-type: none"> • Receiving methotrexate for sarcoidosis may reduce the likelihood of developing an infection compared to azathioprine 	<ul style="list-style-type: none"> • Future studies should establish whether methotrexate can be used as first line treatment in sarcoidosis
VI. Risk and predictors of heart failure in sarcoidosis	<ul style="list-style-type: none"> • Heart failure risks are higher in sarcoidosis compared to the general population, particularly during the first two years after sarcoidosis diagnosis • In most patients, heart failure is associated with a history of ischemic heart disease, but cardiac sarcoidosis may also play a significant role in heart failure etiology • Early detection and control of comorbidity (e.g., diabetes) and cardiac sarcoidosis may help prevent heart failure 	<ul style="list-style-type: none"> • Heart failure risks are higher in individuals with than those without sarcoidosis • Sarcoidosis may predispose to heart failure directly or more commonly through ischemic (atherosclerotic) heart disease • My physician may propose strategies to identify and treat sarcoidosis manifestations and comorbid diseases to reduce the chances that my heart is affected by sarcoidosis 	<ul style="list-style-type: none"> • Clinical studies detailing the clinical characteristics and phenotypes of heart failure in sarcoidosis are needed • Primary and secondary prevention strategies should be evaluated in future studies

9 ACKNOWLEDGEMENTS

I am sincerely thankful for having the following individuals by my side supporting this long educational journey that soon comes to an end:

Elizabeth Arkema, my awesome main supervisor, for giving me countless opportunities, teaching me epidemiological and statistical methods, helping me navigate the research premise, giving me space to evolve as an independent researcher, resolving the small issues that arose throughout, giving useful advice, and being approachable, open-minded, sincere, and fun to work with. I could not have asked for a better supervisor!

Johan Askling, Johan Grunewald, and Daniela Di Giuseppe, my brilliant co-supervisors, who believed in me from the beginning, supported my progress in every way, and encouraged me to become an (almost) independent researcher. Thank you for always providing timely and useful feedback on my manuscripts sharing your long experience as well as thoughtful advice.

Susanna Kullberg and Anders Eklund, my precious collaborators, who shared their deep clinical knowledge of sarcoidosis and their experiences with patients and research and for always being enthusiastic about my work, giving thoughtful comments to my manuscripts and praising me for my efforts.

Kyriaki Kosidou, my mentor, for being friendly, providing me with uplifting advice for the present and future, and for trusting and praising my efforts to do become a better researcher.

Julia Simard, my unofficial mentor, who trusted me at the very beginnings of my research career, for giving me opportunities, helping me improve my understanding of register-based research, methods, and to gain confidence in this highly competitive field. Your dedication to research was and is an inspiration!

Sofia Gernaat, Marina Dehara, Joshua Entrop, and Laura Köcher, current and former members of the Arkema team, whose presence made this journey a little bit more bearable by sharing our frustrations with the research process and the world around us but also our achievements and positive moments. I will miss our conversations.

Thomas Frisell, Marie Holmqvist, Bénédicte Delcoigne, Hannah Bower, Gustaf Bruze, Martin Neovius, Andrei Barbulescu, Matilda Morin, Ola Olén, Hannes Hagström, Mattias Norrbäck, Björn Pasternak, Kristin Waldenlind, Yun-Han Wang, Weng Ian Che, Helga Westerlind, Viktor Wintzell, Hjalmar Wadström, John Svensson, Carl-Martin Pethrus, Peter Ueda, Peter Alping, Michael Sachs, Thomas Bergman, Majd Bairkdar, Anton Öberg Sysojev, and Valérie Leclair, current and former official and unofficial members of the Askling research group, for their useful feedback on my presentations during the group meeting and the scientific (and other) exchange during other activities.

Renata Zelic, Huiling Xu, Kelsi Smith, Katarina Remaeus, and other doctoral students and postdoctoral researchers at KEP and CPE for openly sharing our thoughts, frustrations, and knowledge during lunch time, either related to our work or life in general. I have enjoyed and learned a lot from our discussions.

Michael Fored, Director of doctoral studies at the Department of Medicine, Solna, Helena Nord, Pernilla Appelquist, Eva Willis, and Monica Rundgren, current and past administrators at the Clinical Epidemiology Division, as well as Samir Hussin and Farzad Hedayaty, our IT support team, without the support of whom none of this work would ever be possible, for always being available and willing to promptly resolve small or bigger issues.

All sarcoidosis-interested colleagues and collaborators at Karolinska Institutet and Karolinska Universitetssjukhuset for sharing their clinical experience, research, ideas, and feedback on current and future projects during our monthly cross-field lunch meetings.

Alicia Nevriana, Anna Meyer, Elena Raffetti, Le Zhang, Melody Almroth, Elisa Longinetti, and Alva Wallas, former and current Student Representatives at the Doctoral Program in Epidemiology at Karolinska Institutet, as well as Anita Berglund, director of the Program, and Johanna Bergman, administrator of the Program. I really enjoyed working with you organizing fun and educational activities for our fellow students and colleagues. I hope we will get to collaborate again in the future.

Karolinska Institutet, Karolinska Universitetssjukhuset, and Karolinska Institutet Universitetsbiblioteket for providing top-notch facilities, services, and support enabling this time to run as smooth as possible.

Patients, physicians, and data collectors and managers both at Karolinska Universitetssjukhuset, the regional and the national level who made large-scale research on sarcoidosis possible. I hope this is just the start of a brighter future for epidemiologic research on sarcoidosis.

All funders and supporters of the research projects that I have worked with: Hjärt-Lungfonden, Svenska Läkaresällskapet, Vetenskapsrådet, Karolinska Institutets Stiftelser och Fonder, and Region Stockholm for entrusting us funds from taxes and donations to conduct valuable research on sarcoidosis.

Pandora and Ioannis, my loving parents, for always believing in me and for encouraging and supporting this journey for so long and in every way possible. I am indebted to you for life. Ευχαριστώ από καρδιάς!

Theodoros and Charalambos for their warmheartedly support and concern despite the long distance that keeps us apart all these years.

Alicia, my dearest friend, for being by my side for more than six years supporting this journey, cheering for me, sharing thoughts on work and private matters, dreams, precious moments, and experiences. I could not have asked for a better support team. I hope the best for you wherever life takes you.

Oscar J., my inspiring friend, for leading the way and for all the support and the fun we had during these past several years. Best of luck in the new environment!

Gerson for unconditionally welcoming us to his life and home. I hope to see you guys soon.

Antony, my bright friend, for being an inspiration and my source of advice throughout this time. I have no doubts you will be successful whatever you choose to do in the future. I hope we will get to meet in person soon.

Ying, Christiane, Åsa, Petros, Kyriakos, Kyriakos, Vasilis, Christos, and all other friends of mine in Sweden, the UK, Cyprus, and Greece, for spending time together (usually with food involved), sharing happy memories and dreams for the future. Cannot wait to see you all again!

10 REFERENCES

1. Statement on sarcoidosis. *Am J Respir Crit Care Med*. 1999;160:736–55.
2. Grunewald J, Grutters JC, Arkema E V., Saketkoo LA, Moller DR, Müller-Quernheim J. Sarcoidosis. *Nat Rev Dis Prim*. 2019;5:45.
3. Spagnolo P, Rossi G, Trisolini R, Sverzellati N, Baughman RP, Wells AU. Pulmonary sarcoidosis. *Lancet Respir Med*. 2018;6:389–402.
4. Karakaya B, Kaiser Y, van Moorsel C, Grunewald J. Löfgren's syndrome: diagnosis, management, and disease pathogenesis. *Semin Respir Crit Care Med*. 2017;38:463–76.
5. Iriarte A, Rubio-Rivas M, Villalba N, Corbella X, Mañá J. Clinical features and outcomes of asymptomatic pulmonary sarcoidosis. A comparative cohort study. *Respir Med*. 2020;169:105998.
6. Hillerdal G, Nöu E, Osterman K, Schmekel B. Sarcoidosis: epidemiology and prognosis. A 15-year European study. *Am Rev Respir Dis*. 1984;130:29–32.
7. Sharma OP. Sarcoidosis in history. In: Mitchell DN, Moller DR, Spiro SG, Wells A, editors. *Sarcoidosis*. Boca Radon, FL: CRC Press; 2012. p. 3–11.
8. Spagnolo P. Sarcoidosis: a Critical Review of History and Milestones. *Clin Rev Allergy Immunol*. 2015;49:1–5.
9. Löfgren S. Primary Pulmonary Sarcoidosis. *Acta Med Scand*. 2009;145:424–31.
10. Rossi G, Cavazza A, Colby T V. Pathology of Sarcoidosis. *Clin Rev Allergy Immunol*. 2015;49:36–44.
11. Corrin B, Nicholson AG, Dewar A. Pathology. In: Mitchell DN, Moller DR, Spiro SG, Wells A, editors. *Sarcoidosis*. Boca Radon, FL: CRC Press; 2012. p. 41–2.
12. Broos CE, van Nimwegen M, Hoogsteden HC, Hendriks RW, Kool M, van den Blink B. Granuloma formation in pulmonary sarcoidosis. *Front Immunol*. 2013;4.
13. Wahlström J, Dengjel J, Persson B, Duyar H, Rammensee H-G, Stevanović S, et al. Identification of HLA-DR-bound peptides presented by human bronchoalveolar lavage cells in sarcoidosis. *J Clin Invest*. American Society for Clinical Investigation; 2007;117:3576–82.
14. Beijer E, Veltkamp M, Meek B, Moller D. Etiology and immunopathogenesis of sarcoidosis: novel insights. *Semin Respir Crit Care Med*. 2017;38:404–16.
15. Kaiser Y, Eklund A, Grunewald J. Moving target: Shifting the focus to pulmonary sarcoidosis as an autoimmune spectrum disorder. *Eur Respir J*. 2019;54:1–14.
16. Grunewald J, Eklund A. Löfgren's syndrome: human leukocyte antigen strongly influences the disease course. *Am J Respir Crit Care Med*. 2009;179:307–12.
17. Birnie DH. Cardiac Sarcoidosis. *Semin Respir Crit Care Med*. 2020;41:626–40.
18. Okasha O, Kazmirczak F, Chen KA, Farzaneh-Far A, Shenoy C. Myocardial Involvement in Patients With Histologically Diagnosed Cardiac Sarcoidosis: A Systematic Review and Meta-Analysis of Gross Pathological Images From Autopsy or Cardiac Transplantation Cases. *J Am Heart Assoc*. 2019;8:e011253.

19. Birnie DH, Nery PB, Ha AC, Beanlands RSB. Cardiac Sarcoidosis. *J Am Coll Cardiol*. 2016;68:411–21.
20. Okada DR, Smith J, Derakhshan A, Gowani Z, Misra S, Berger RD, et al. Ventricular arrhythmias in cardiac sarcoidosis. *Circulation*. 2018;138:1253–64.
21. Ekström K, Lehtonen J, Nordenswan HK, Mäyränpää MI, Räisänen-Sokolowski A, Kandolin R, et al. Sudden death in cardiac sarcoidosis: An analysis of nationwide clinical and cause-of-death registries. *Eur Heart J*. 2019;40:3121–8.
22. Birnie DH, Sauer WH, Bogun F, Cooper JM, Culver DA, Duvernoy CS, et al. HRS expert consensus statement on the diagnosis and management of arrhythmias associated with cardiac sarcoidosis. *Hear Rhythm*. 2014;11:1304–23.
23. Kebed KY, Carter S V., Flatley E, Ward RP, Moss JD, Appelbaum DE, et al. Prevalence of newly diagnosed sarcoidosis in patients with ventricular arrhythmias: a cardiac magnetic resonance and 18F-FDG cardiac PET study. *Int J Cardiovasc Imaging*. 2020;[In press].
24. Franke KB, Marshall H, Kennewell P, Pham H-D, Tully PJ, Rattanakositt T, et al. Risk and predictors of sudden death in cardiac sarcoidosis: A systematic review and meta-analysis. *Int J Cardiol*. 2020;[In press].
25. Judson MA. The diagnosis of sarcoidosis. *Curr Opin Pulm Med*. 2019;25:484–96.
26. Węclawek M, Ziora D, Jastrzębski D. Imaging methods for pulmonary sarcoidosis. *Adv Respir Med*. 2020;88:18–26.
27. Crouser ED, Maier LA, Wilson KC, Bonham CA, Morgenthau AS, Patterson KC, et al. Diagnosis and Detection of Sarcoidosis. An Official American Thoracic Society Clinical Practice Guideline. *Am J Respir Crit Care Med*. 2020;201:e26–51.
28. Nunes H, Uzunhan Y, Gille T, Lamberto C, Valeyre D, Brillet P-Y. Imaging of sarcoidosis of the airways and lung parenchyma and correlation with lung function. *Eur Respir J*. 2012;40:750–65.
29. Aryal S, Nathan SD. Contemporary optimized practice in the management of pulmonary sarcoidosis. *Ther Adv Respir Dis*. 2019;13:1–15.
30. Baughman RP, Sparkman BK, Lower EE. Six-Minute Walk Test and Health Status Assessment in Sarcoidosis. *Chest*. 2007;132:207–13.
31. Brice EAW, Friedlander W, Bateman ED, Kirsch BE. Serum Angiotensin-Converting Enzyme Activity, Concentration, and Specific Activity in Granulomatous Interstitial Lung Disease, Tuberculosis, and COPD. *Chest*. 1995;107:706–10.
32. Miyoshi S, Hamada H, Kadowaki T, Hamaguchi N, Ito R, Irifune K, et al. Comparative Evaluation of Serum Markers in Pulmonary Sarcoidosis. *Chest*. 2010;137:1391–7.
33. Bennett D, Cameli P, Lanzarone N, Carobene L, Bianchi N, Fui A, et al. Chitotriosidase: a biomarker of activity and severity in patients with sarcoidosis. *Respir Res*. 2020;21:6.
34. Zhou Y, Zhang Y, Zhao M, Li Q, Li H. sIL-2R levels predict the spontaneous remission in sarcoidosis. *Respir Med*. 2020;171:106115.
35. Thillai M, Chang W, Chaudhuri N, Forrest I, Ho L-P, Lines S, et al. Sarcoidosis in the UK: insights from British Thoracic Society registry data. *BMJ Open Respir Res*. 2019;6:e000357.

36. d'Alessandro M, Bergantini L, Perrone A, Cameli P, Cameli M, Prasse A, et al. Serial investigation of Angiotensin-Converting Enzyme in sarcoidosis patients treated with Angiotensin-Converting Enzyme Inhibitor. *Eur J Intern Med.* 2020;78:58–62.
37. Linke M, Pham HTT, Katholnig K, Schnöller T, Miller A, Demel F, et al. Chronic signaling via the metabolic checkpoint kinase mTORC1 induces macrophage granuloma formation and marks sarcoidosis progression. *Nat Immunol.* 2017;18:293–302.
38. Pacheco Y, Lim CX, Weichhart T, Valeyre D, Bentaher A, Calender A. Sarcoidosis and the mTOR, Rac1, and Autophagy Triad. *Trends Immunol.* 2020;41:286–99.
39. Kraaijvanger R, Janssen Bonás M, Vorselaars A, Veltkamp M. Biomarkers in the Diagnosis and Prognosis of Sarcoidosis: Current Use and Future Prospects. *Front Immunol.* 2020;11:1–17.
40. Wasfi YS, Rose CS, Murphy JR, Silveira LJ, Grutters JC, Inoue Y, et al. A New Tool To Assess Sarcoidosis Severity. *Chest.* 2006;129:1234–45.
41. Yeung H, Farber S, Birnbaum BK, Dunham J, Ogdie A, Patterson KC, et al. Reliability and Validity of Cutaneous Sarcoidosis Outcome Instruments Among Dermatologists, Pulmonologists, and Rheumatologists. *JAMA Dermatology.* 2015;151:1317.
42. Baughman RP, Tillinger M, Qin Y, Sweiss N, Lower EE. A composite score to assess treatment response in pulmonary sarcoidosis: The sarcoidosis treatment Score (STS). *Sarcoidosis, Vasc Diffus Lung Dis.* 2019;36:86–8.
43. Lopes MC, Amadeu TP, Ribeiro-Alves M, Costa CH da, Silva BRA, Rodrigues LS, et al. Defining prognosis in sarcoidosis. *Medicine (Baltimore).* 2020;99:e23100.
44. Scadding JG. Prognosis of intrathoracic sarcoidosis in England. A review of 136 cases after five years' observation. *Br Med J.* 1961;2:1165–72.
45. Levy A, Hamzeh N, Maier LA. Is it time to scrap Scadding and adopt computed tomography for initial evaluation of sarcoidosis? *F1000Research.* 2018;7:600.
46. Culver DA, Baughman RP. It's time to evolve from Scadding: phenotyping sarcoidosis. *Eur Respir J.* 2018;51:1800050.
47. Rahaghi FF, Baughman RP, Saketkoo LA, Sweiss NJ, Barney JB, Birring SS, et al. Delphi consensus recommendations for a treatment algorithm in pulmonary sarcoidosis. *Eur Respir Rev.* 2020;29:190146.
48. Schutt AC, Bullington WM, Judson MA. Pharmacotherapy for pulmonary sarcoidosis: A delphi consensus study. *Respir Med.* 2010;104:717–23.
49. Lingner H, Buhr-Schinner H, Hummel S, van der Meyden J, Grosshennig A, Nowik D, et al. Short-Term Effects of a Multimodal 3-Week Inpatient Pulmonary Rehabilitation Programme for Patients with Sarcoidosis: The ProKaSaRe Study. *Respiration.* 2018;95:343–53.
50. James WE, Baughman R. Treatment of sarcoidosis: grading the evidence. *Expert Rev Clin Pharmacol.* 2018;11:677–87.
51. Baughman RP, Judson MA, Wells A. The indications for the treatment of sarcoidosis: Wells law. *Sarcoidosis, Vasc Diffus Lung Dis.* 2017;34:280–2.

52. Nunes H, Jeny F, Bouvry D, Uzunhan Y, Valeyre D. Indications for treatment of sarcoidosis. *Curr Opin Pulm Med*. 2019;25:505–18.
53. FSR. Sarcoidosis treatment guidelines. Chicago, IL: Foundation for Sarcoidosis Research; 2018.
54. Valeyre D, Jeny F, Nunes H. Current medical therapy for sarcoidosis. *Semin Respir Crit Care Med*. 2017;38:523–31.
55. Svensk Lungmedicinsk Förening. Vårdprogram för sarkoidos [Sarcoidosis care recommendations]. Göteborg: Mediahuset AB; 2018.
56. Gibson GJ, Prescott RJ, Muers MF, Middleton WG, Mitchell DN, Connolly CK, et al. British Thoracic Society Sarcoidosis study: Effects of long term corticosteroid treatment. *Thorax*. 1996;51:238–47.
57. Judson MA. Developing better drugs for pulmonary sarcoidosis: determining indications for treatment and endpoints to assess therapy based on patient and clinician concerns. *F1000Research*. 2019;8:2149.
58. Rossides M, Kullberg S, Eklund A, Grunewald J, Arkema E V. Sarcoidosis diagnosis and treatment in Sweden: A register-based assessment of variations by region and calendar period. *Respir Med*. 2020;161:105846.
59. Pande A, Culver DA. Knowing when to use steroids, immunosuppressants or biologics for the treatment of sarcoidosis. *Expert Rev Respir Med*. 2020;14:285–98.
60. Baughman RP, Winget DB, Lower EE. Methotrexate is steroid sparing in acute sarcoidosis: results of a double blind, randomized trial. *Sarcoidosis, Vasc Diffus Lung Dis*. 2000;17:60–6.
61. Vorselaars ADM, Wuyts WA, Vorselaars VMM, Zanen P, Deneer VHM, Velthuis M, et al. Methotrexate vs azathioprine in second-line therapy of sarcoidosis. *Chest*. 2013;144:805–12.
62. Saketkoo LA, Baughman RP. Biologic therapies in the treatment of sarcoidosis. *Expert Rev Clin Immunol*. 2016;12:817–25.
63. Goldman C, Judson MA. Corticosteroid refractory sarcoidosis. *Respir Med*. 2020;171:106081.
64. Gerke AK, Tang F, Pendergast J, Cavanaugh JE, Polgreen PM. The high frequency of healthcare use in patients one year prior to a sarcoidosis diagnosis. *Sarcoidosis, Vasc Diffus Lung Dis*. 2014;31:256–61.
65. Arkema E V, Eklund A, Grunewald J, Bruze G. Work ability before and after sarcoidosis diagnosis in Sweden. *Respir Med*. 2018;144:S7–12.
66. Gade Sikjær M, Hilberg O, Ibsen R, Bock K, Løkke A. Direct and indirect economic and health consequences related to sarcoidosis in Denmark: a national register-based study. *Respir Med*. 2019;152:7–13.
67. Ma SP, Rogers SL, Hall AJ, Hodgson L, Brennan J, Stawell RJ, et al. Sarcoidosis-related uveitis: clinical presentation, disease course, and rates of systemic disease progression after uveitis diagnosis. *Am J Ophthalmol*. 2019;198:30–6.
68. Valeyre D, Prasse A, Nunes H, Uzunhan Y, Brillet P-YY, Müller-Quernheim J. Sarcoidosis. *Lancet*. 2014;383:1155–67.

69. Takada K, Ina Y, Noda M, Sato T, Yamamoto M, Morishita M. The clinical course and prognosis of patients with severe, moderate or mild sarcoidosis. *J Clin Epidemiol.* 1993;46:359–66.
70. Neville E, Walker AN, James DG. Prognostic Factors Predicting the Outcome of Sarcoidosis: An Analysis of 818 Patients. *QJM.* 1983;52:525–33.
71. Rubio-Rivas M, Corbella X. Clinical phenotypes and prediction of chronicity in sarcoidosis using cluster analysis in a prospective cohort of 694 patients. *Eur J Intern Med.* 2020;77:59–65.
72. Arkema E V, Cozier YC. Epidemiology of sarcoidosis: current findings and future directions. *Ther Adv Chronic Dis.* 2018;9:227–40.
73. Arkema E V., Cozier YC. Sarcoidosis epidemiology: recent estimates of incidence, prevalence and risk factors. *Curr Opin Pulm Med.* 2020;26:527–34.
74. Baughman RP, Field S, Costabel U, Crystal RG, Culver DA, Drent M, et al. Sarcoidosis in America. Analysis based on health care use. *Ann Am Thorac Soc.* 2016;13:1244–52.
75. Cozier YC, Berman JS, Palmer JR, Boggs DA, Serlin DM, Rosenberg L. Sarcoidosis in Black women in the United States. *Chest.* 2011;139:144–50.
76. Arkema E V, Grunewald J, Kullberg S, Eklund A, Askling J. Sarcoidosis incidence and prevalence: a nationwide register-based assessment in Sweden. *Eur Respir J.* 2016;48:1690–9.
77. ACCESS Research Group. Design of A Case Control Etiologic Study of Sarcoidosis (ACCESS). *J Clin Epidemiol.* 1999;52:1173–86.
78. Ungprasert P, Carmona EM, Utz JP, Ryu JH, Crowson CS, Matteson EL. Epidemiology of sarcoidosis 1946–2013: a population-based study. *Mayo Clin Proc.* 2016;91:183–8.
79. Cozier YC, Coogan PF, Govender P, Berman JS, Palmer JR, Rosenberg L. Obesity and weight gain in relation to incidence of sarcoidosis in US Black women. *Chest.* 2015;147:1086–93.
80. Dumas O, Abramovitz L, Wiley AS, Cozier YC, Camargo CA. Epidemiology of sarcoidosis in a prospective cohort study of U.S. women. *Ann Am Thorac Soc.* 2016;13:67–71.
81. Wu C-H, Chung P-I, Wu C-Y, Chen Y-T, Chiu Y-W, Chang Y-T, et al. Comorbid autoimmune diseases in patients with sarcoidosis: A nationwide case-control study in Taiwan. *J Dermatol.* 2017;44:423–30.
82. Wang S, Chung C, Huang T, Tsai W, Peng C, Huang K, et al. Bidirectional association between tuberculosis and sarcoidosis. *Respirology.* 2019;24:467–74.
83. Gribbin J, Hubbard RB, Le Jeune I, Smith CJP, West J, Tata LJ. Incidence and mortality of idiopathic pulmonary fibrosis and sarcoidosis in the UK. *Thorax.* 2006;61:980–5.
84. Rossides M, Kullberg S, Askling J, Eklund A, Grunewald J, Arkema E V. Sarcoidosis mortality in Sweden: a population-based cohort study. *Eur Respir J.* 2018;51:1701815.
85. Rossides M, Grunewald J, Eklund A, Kullberg S, Di Giuseppe D, Askling J, et al. Familial aggregation and heritability of sarcoidosis: a Swedish nested case–control study. *Eur Respir J.* 2018;52:1800385.

86. Nathan N, Sileo C, Calender A, Pacheco Y, Rosental P-A, Cavalin C, et al. Paediatric sarcoidosis. *Paediatr Respir Rev*. 2019;29:53–9.
87. Cozier YC, Berman JS, Palmer JR, Boggs DA, Wise LA, Rosenberg L. Reproductive and Hormonal Factors in Relation to Incidence of Sarcoidosis in US Black Women: The Black Women’s Health Study. *Am J Epidemiol*. 2012;176:635–41.
88. Gillman A, Steinfort C. Sarcoidosis in Australia. *Intern Med J*. 2007;37:356–9.
89. Fidler LM, Balter M, Fisher JH, To T, Stanbrook MB, Gershon A. Epidemiology and health outcomes of sarcoidosis in a universal healthcare population: a cohort study. *Eur Respir J*. 2019;54:1900444.
90. Byg K-E, Milman N, Hansen S. Sarcoidosis in Denmark 1980-1994. A registry-based incidence study comprising 5536 patients. *Sarcoidosis, Vasc Diffus Lung Dis*. 2003;20:46–52.
91. Pietinalho A. Sarcoidosis in Finland and Hokkaido, Japan: A study of two genetically different populations. University of Helsinki; 2000.
92. Duchemann B, Annesi-Maesano I, Jacobe de Naurois C, Sanyal S, Brillet P-Y, Brauner M, et al. Prevalence and incidence of interstitial lung diseases in a multi-ethnic county of Greater Paris. *Eur Respir J*. 2017;50:1602419.
93. Karakatsani A, Papakosta D, Rapti A, Antoniou KM, Dimadi M, Markopoulou A, et al. Epidemiology of interstitial lung diseases in Greece. *Respir Med*. 2009;103:1122–9.
94. Beghè D, Dall’Asta L, Garavelli C, Pastorelli AA, Muscarella M, Saccani G, et al. Sarcoidosis in an Italian province. Prevalence and environmental risk factors. *PLoS One*. 2017;12:e0176859.
95. Morimoto T, Azuma A, Abe S, Usuki J, Kudoh S, Sugisaki K, et al. Epidemiology of sarcoidosis in Japan. *Eur Respir J*. 2008;31:372–9.
96. Kowalska M, Niewiadomska E, Zejda JE. Epidemiology of sarcoidosis recorded in 2006-2010 in the silesian voivodeship on the basis of routine medical reporting. *Ann Agric Environ Med*. 2014;21:55–8.
97. Park JE, Kim YS, Kang MJ, Kim CJ, Han CH, Lee SM, et al. Prevalence, incidence, and mortality of sarcoidosis in Korea, 2003–2015: A nationwide population-based study. *Respir Med*. 2018;144:S28–34.
98. Deubelbeiss U, Gemperli A, Schindler C, Baty F, Brutsche MH. Prevalence of sarcoidosis in Switzerland is associated with environmental factors. *Eur Respir J*. 2010;35:1088–97.
99. Kajdasz DK, Judson MA, Mohr LC, Lackland DT. Geographic variation in sarcoidosis in South Carolina: its Relation to socioeconomic status and health care indicators. *Am J Epidemiol*. 1999;150:271–8.
100. Bauer HJ, Wijkström S. The Prevalence of Pulmonary Sarcoidosis in Swedish Mass Radiography Surveys. *Acta Med Scand*. 1964;176:112–4.
101. Ungprasert P, Crowson CS, Matteson EL. Seasonal variation in incidence of sarcoidosis: a population-based study, 1976–2013. *Thorax*. 2016;71:1164–6.
102. Demirkok SS, Basaranoglu M, Akbilgic O. Seasonal variation of the onset of presentations in stage 1 sarcoidosis. *Int J Clin Pract*. 2006;60:1443–50.

103. Gerke AK, Tangh F, Yang M, Cavanaugh JE, Polgreen PM. An analysis of seasonality of sarcoidosis in the United States veteran population: 2000-2007. *Sarcoidosis, Vasc Diffus Lung Dis.* 2012;29:155–8.
104. Wiman LG. Familial occurrence of sarcoidosis. *Scand J Respir Dis Suppl.* 1972;80:115–9.
105. Familial associations in sarcoidosis. A report to the Research Committee of the British Thoracic and Tuberculosis Association. *Tubercle.* 1973;54:87–98.
106. Headings VE, Weston D, Young RC, Hackney RL. Familial sarcoidosis with multiple occurrences in eleven families: a possible mechanism of inheritance. *Ann N Y Acad Sci.* 1976;278:377–85.
107. Brennan NJ, Crean P, Long JP, Fitzgerald MX. High prevalence of familial sarcoidosis in an Irish population. *Thorax.* 1984;39:14–8.
108. Fité E, Alsina JM, Antó JM, Morera J. Sarcoidosis: family contact study. *Respiration.* 1998;65:34–9.
109. McGrath DS, Daniil Z, Foley P, du Bois JL, Lympny PA, Cullinan P, et al. Epidemiology of familial sarcoidosis in the UK. *Thorax.* 2000;55:751–4.
110. Rybicki BA, Kirkey KL, Major M, Maliarik MJ, Popovich Jr. J, Chase GA, et al. Familial risk ratio of sarcoidosis in African-American sibs and parents. *Am J Epidemiol.* 2001;153:188–93.
111. Rybicki BA, Iannuzzi MC, Frederick MM, Thompson BW, Rossman MD, Bresnitz EA, et al. Familial aggregation of sarcoidosis. A Case-Control Etiologic Study of Sarcoidosis (ACCESS). *Am J Respir Crit Care Med.* 2001;164:2085–91.
112. Falconer DS, Mackay TF. *Introduction to Quantitative Genetics.* 4th ed. Harlow, UK: Pearson; 1996.
113. Sverrild A, Backer V, Kyvik KO, Kaprio J, Milman N, Svendsen CB, et al. Heredity in sarcoidosis: a registry-based twin study. *Thorax.* 2008;63:894–6.
114. Schürmann M, Reichel P, Müller-Myhsok B, Schlaak M, Müller-Quernheim J, Schwinger E. Results from a Genome-wide Search for Predisposing Genes in Sarcoidosis. *Am J Respir Crit Care Med.* 2001;164:840–6.
115. Iannuzzi MC, Iyengar SK, Gray-McGuire C, Elston RC, Baughman RP, Donohue JF, et al. Genome-wide search for sarcoidosis susceptibility genes in African Americans. *Genes Immun.* 2005;6:509–18.
116. Rivera N V, Ronninger M, Shchetynsky K, Franke A, Nöthen MM, Müller-Quernheim J, et al. High-density genetic mapping identifies new susceptibility variants in sarcoidosis phenotypes and shows genomic-driven phenotypic differences. *Am J Respir Crit Care Med.* 2016;193:1008–22.
117. Franke A, Fischer A, Nothnagel M, Becker C, Grabe N, Till A, et al. Genome-wide association analysis in sarcoidosis and Crohn's disease unravels a common susceptibility locus on 10p12.2. *Gastroenterology.* 2008;135:1207–15.

118. Hofmann S, Franke A, Fischer A, Jacobs G, Nothnagel M, Gaede KI, et al. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet.* 2008;40:1103–6.
119. Sato H, Woodhead FA, Ahmad T, Grutters JC, Spagnolo P, van den Bosch JM, et al. Sarcoidosis HLA class II genotyping distinguishes differences of clinical phenotype across ethnic groups. *Hum Mol Genet.* 2010;19:4100–11.
120. Meguro A, Ishihara M, Petrek M, Yamamoto K, Takeuchi M, Mrazek F, et al. Genetic control of CCL24, POR, and IL23R contributes to the pathogenesis of sarcoidosis. *Commun Biol.* 2020;3:465.
121. Garman L, Montgomery CG, Rivera N V. Recent advances in sarcoidosis genomics: epigenetics, gene expression, and gene by environment ($G \times E$) interaction studies. *Curr Opin Pulm Med.* 2020;26:544–53.
122. Calender A, Rollat Farnier PA, Buisson A, Pinson S, Bentaher A, Lebecque S, et al. Whole exome sequencing in three families segregating a pediatric case of sarcoidosis. *BMC Med Genomics.* 2018;11:23.
123. Besnard V, Calender A, Bouvry D, Pacheco Y, Chapelon-Abrie C, Jeny F, et al. G908R NOD2 variant in a family with sarcoidosis. *Respir Res.* 2018;19:44.
124. Calender A, Weichhart T, Valeyre D, Pacheco Y. Current Insights in Genetics of Sarcoidosis: Functional and Clinical Impacts. *J Clin Med.* 2020;9:2633.
125. Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A, et al. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat Genet.* 2005;37:357–64.
126. Fischer A, Ellinghaus D, Nutsua M, Hofmann S, Montgomery CG, Iannuzzi MC, et al. Identification of Immune-Relevant Factors Conferring Sarcoidosis Genetic Risk. *Am J Respir Crit Care Med.* 2015;192:727–36.
127. Pacheco Y, Calender A, Israël-Biet D, Roy P, Lebecque S, Cottin V, et al. Familial vs. sporadic sarcoidosis: BTNL2 polymorphisms, clinical presentations, and outcomes in a French cohort. *Orphanet J Rare Dis.* 2016;11:165.
128. Wolin A, Lahtela EL, Anttila V, Petrek M, Grunewald J, van Moorsel CHM, et al. SNP Variants in Major Histocompatibility Complex Are Associated with Sarcoidosis Susceptibility—A Joint Analysis in Four European Populations. *Front Immunol.* 2017;8:422.
129. Yang I V., Konigsberg I, MacPhail K, Li L, Davidson EJ, Mroz PM, et al. DNA methylation changes in lung immune cells are associated with granulomatous lung disease. *Am J Respir Cell Mol Biol.* 2019;60:96–105.
130. Rivera N V., Patasova K, Kullberg S, Diaz-Gallo LM, Iseda T, Bengtsson C, et al. A Gene–Environment Interaction Between Smoking and Gene polymorphisms Provides a High Risk of Two Subgroups of Sarcoidosis. *Sci Rep.* 2019;9:18633.
131. Chen Y, Adrianto I, Iannuzzi MC, Garman L, Montgomery CG, Rybicki BA, et al. Extended methods for gene–environment-wide interaction scans in studies of admixed individuals with varying degrees of relationships. *Genet Epidemiol.* 2019;43:414–26.
132. Saidha S, Sotirchos ES, Eckstein C. Etiology of sarcoidosis: does infection play a role? *Yale J Biol Med.* 2012;85:133–41.

133. Esteves T, Aparicio G, Garcia-Patos V. Is there any association between sarcoidosis and infectious agents?: a systematic review and meta-analysis. *BMC Pulm Med*. 2016;16:165.
134. Gupta D, Agarwal R, Aggarwal AN, Jindal SK. Molecular evidence for the role of mycobacteria in sarcoidosis: a meta-analysis. *Eur Respir J*. 2007;30:508–16.
135. Yamada T, Eishi Y, Ikeda S, Ishige I, Suzuki T, Takemura T, et al. In situ localization of *Propionibacterium acnes* DNA in lymph nodes from sarcoidosis patients by signal amplification with catalysed reporter deposition. *J Pathol*. 2002;198:541–7.
136. Svendsen CB, Milman N, Høier-Madsen M, Dziegiel MH, Krogfelt KA. Determination of rickettsial and antinuclear antibodies in Danish patients with sarcoidosis. *Clin Respir J*. 2008;2:202–7.
137. Hofland RW, Thijsen SF, Bouwman J, van der Wel M, Bossink AW. Sarcoidosis and purified protein derivative reactivity. *Sarcoidosis, Vasc Diffus Lung Dis*. 2014;31:142–8.
138. Lebbe C, Agbalika F, Flageul B, Pellet C, Rybojad M, Cordoliani F, et al. No evidence for a role of human herpesvirus type 8 in sarcoidosis: molecular and serological analysis. *Br J Dermatol*. 1999;141:492–6.
139. Terčelj M, Stopinšek S, Ihan A, Salobir B, Simčič S, Rylander R. Fungal exposure and low levels of IL-10 in patients with sarcoidosis. *Pulm Med*. 2014;2014:1–4.
140. Suchankova M, Paulovicova E, Paulovicova L, Majer I, Tedlova E, Novosadova H, et al. Increased antifungal antibodies in bronchoalveolar lavage fluid and serum in pulmonary sarcoidosis. *Scand J Immunol*. 2015;81:259–64.
141. Clarke EL, Lauder AP, Hofstaedter CE, Hwang Y, Fitzgerald AS, Imai I, et al. Microbial lineages in sarcoidosis. A metagenomic analysis tailored for low-microbial content samples. *Am J Respir Crit Care Med*. 2018;197:225–34.
142. Garzoni C, Brugger SD, Qi W, Wasmer S, Cusini A, Dumont P, et al. Microbial communities in the respiratory tract of patients with interstitial lung disease. *Thorax*. 2013;68:1150–6.
143. Zimmermann A, Knecht H, Häsler R, Zissel G, Gaede KI, Hofmann S, et al. *Atopobium* and *Fusobacterium* as novel candidates for sarcoidosis-associated microbiota. *Eur Respir J*. 2017;50:1600746.
144. Judson MA. Environmental Risk Factors for Sarcoidosis. *Front Immunol*. 2020;11:1–10.
145. Rotsinger JE, Celada LJ, Polosukhin V V, Atkinson JB, Drake WP. Molecular Analysis of Sarcoidosis Granulomas Reveals Antimicrobial Targets. *Am J Respir Cell Mol Biol*. 2016;55:128–34.
146. Chen ES, Wahlström J, Song Z, Willett MH, Wikén M, Yung RC, et al. T Cell Responses to Mycobacterial Catalase-Peroxidase Profile a Pathogenic Antigen in Systemic Sarcoidosis. *J Immunol*. 2008;181:8784–96.
147. Milman N, Lisby G, Friis S, Kemp L. Prolonged culture for mycobacteria in mediastinal lymph nodes from patients with pulmonary sarcoidosis. A negative study. *Sarcoidosis Vasc Diffus Lung Dis*. 2004/05/07. 2004;21:25–8.

148. Suzuki Y, Uchida K, Takemura T, Sekine M, Tamura T, Furukawa A, et al. Propionibacterium acnes-derived insoluble immune complexes in sinus macrophages of lymph nodes affected by sarcoidosis. *PLoS One*. 2018;13:e0192408.
149. Yorozu P, Furukawa A, Uchida K, Akashi T, Kakegawa T, Ogawa T, et al. Propionibacterium acnes catalase induces increased Th1 immune response in sarcoidosis patients. *Respir Investig*. 2015;53:161–9.
150. Derler AM, Eisendle K, Baltaci M, Obermoser G, Zelger B. High prevalence of “Borrelia-like” organisms in skin biopsies of sarcoidosis patients from Western Austria. *J Cutan Pathol*. 2009;36:1262–8.
151. Ishihara M, Ishida T, Isogai E, Kimura K, Oritsu M, Matsui Y, et al. Detection of antibodies to Borrelia species among patients with confirmed sarcoidosis in a region where Lyme disease is nonendemic. *Graefes Arch Clin Exp Ophthalmol*. 1996;234:770–3.
152. Di Alberti L, Piatteli A, Artese L, Favia G, Porter SR, Scully CM, et al. Human herpesvirus 8 variants in sarcoid tissues. *Lancet*. 1997;350:1655–61.
153. Terčelj M, Salobir B, Harlander M, Rylander R. Fungal exposure in homes of patients with sarcoidosis - an environmental exposure study. *Environ Heal*. 2011;10:8.
154. Drake WP, Oswald-Richter K, Richmond BW, Isom J, Burke VE, Algood H, et al. Oral antimycobacterial therapy in chronic cutaneous sarcoidosis. *JAMA Dermatology*. 2013;149:1040–9.
155. Phase II Investigation of Antimycobacterial Therapy on Progressive, Pulmonary Sarcoidosis [Internet]. *ClinicalTrials.gov*. 2020 [cited Aug 10, 2020]. Available from: <https://clinicaltrials.gov/ct2/show/study/NCT02024555>
156. Ishibashi K, Eishi Y, Tahara N, Asakura M, Sakamoto N, Nakamura K, et al. Japanese Antibacterial Drug Management for Cardiac Sarcoidosis (J-ACNES): A multicenter, open-label, randomized, controlled study. *J Arrhythmia*. 2018;34:520–6.
157. Pradhan S, Madke B, Kabra P, Singh A. Anti-inflammatory and immunomodulatory effects of antibiotics and their use in dermatology. *Indian J Dermatol*. 2016;61:469.
158. Nau R, Tauber S. Immunomodulatory Properties of Antibiotics. *Curr Mol Pharmacol*. 2008;1:68–79.
159. Newman KL, Newman LS. Occupational causes of sarcoidosis. *Curr Opin Allergy Clin Immunol*. 2012;12:145–50.
160. Vihlborg P, Bryngelsson I-L, Andersson L, Graff P. Risk of sarcoidosis and seropositive rheumatoid arthritis from occupational silica exposure in Swedish iron foundries: a retrospective cohort study. *BMJ Open*. 2017;7:e016839.
161. Hena KM, Yip J, Jaber N, Goldfarb D, Fullam K, Cleven K, et al. Clinical course of sarcoidosis in World Trade Center-exposed firefighters. *Chest*. 2018;153:114–23.
162. Newman LS, Rose CS, Bresnitz EA, Rossman MD, Barnard J, Frederick M, et al. A Case Control Etiologic Study of Sarcoidosis: environmental and occupational risk factors. *Am J Respir Crit Care Med*. 2004;170:1324–30.

163. Prezant DJ, Dhala A, Goldstein A, Janus D, Ortiz F, Aldrich TK, et al. The Incidence, Prevalence, and Severity of Sarcoidosis in New York City Firefighters. *Chest*. 1999;116:1183–93.
164. Gorham ED, Garland CF, Garland FC, Kaiser K, Travis WD, Centeno JA. Trends and Occupational Associations in Incidence of Hospitalized Pulmonary Sarcoidosis and Other Lung Diseases in Navy Personnel. *Chest*. 2004;126:1431–8.
165. Jonsson E, Järholm B, Andersson M. Silica dust and sarcoidosis in Swedish construction workers. *Occup Med (Chic Ill)*. 2019;69:482–6.
166. Kajdasz D, Lackland DT, Mohr LC, Judson MA. A Current Assessment of Rurally Linked Exposures as Potential Risk Factors for Sarcoidosis. *Ann Epidemiol*. 2001;11:111–7.
167. Gupta D, Singh AD, Agarwal R, Aggarwal AN, Joshi K, Jindal SK. Is tobacco smoking protective for sarcoidosis? A case-control study from North India. *Sarcoidosis, Vasc Diffus Lung Dis*. 2010;27:19–26.
168. Ungprasert P, Crowson CS, Matteson EL. Smoking, obesity and risk of sarcoidosis: A population-based nested case-control study. *Respir Med*. 2016;120:87–90.
169. Carlens C, Hergens M-P, Grunewald J, Ekblom A, Eklund A, Höglund CO, et al. Smoking, use of moist snuff, and risk of chronic inflammatory diseases. *Am J Respir Crit Care Med*. 2010;181:1217–22.
170. Ström KE, Eklund AG. Smoking does not prevent the onset of respiratory failure in sarcoidosis. *Sarcoidosis*. 1993;10:26–8.
171. Margaritopoulos GA, Vasarmidi E, Jacob J, Wells AU, Antoniou KM. Smoking and interstitial lung diseases. *Eur Respir Rev*. 2015;24:428–35.
172. Dumas O, Boggs KM, Cozier YC, Stampfer MJ, Camargo CA. Prospective study of body mass index and risk of sarcoidosis in US women. *Eur Respir J*. 2017;50:1701397.
173. Harpsøe MC, Basit S, Andersson M, Nielsen NM, Frisch M, Wohlfahrt J, et al. Body mass index and risk of autoimmune diseases: a study within the Danish National Birth Cohort. *Int J Epidemiol*. 2014;43:843–55.
174. Guzik TJ, Skiba DS, Touyz RM, Harrison DG. The role of infiltrating immune cells in dysfunctional adipose tissue. *Cardiovasc Res*. 2017;113:1009–23.
175. Wells AU. Sarcoidosis: A benign disease or a culture of neglect? *Respir Med*. 2018;144:S1–2.
176. Jamilloux Y, Maucourt-Boulch D, Kerever S, Gerfaud-Valentin M, Broussolle C, Eb M, et al. Sarcoidosis-related mortality in France: a multiple-cause-of-death analysis. *Eur Respir J*. 2016;48:1700–9.
177. Duncan ME, Goldacre MJ. Mortality trends for tuberculosis and sarcoidosis in English populations, 1979–2008. *Int J Tuberc Lung Dis*. 2012;16:38–42.
178. Swigris JJ, Olson AL, Huie TJ, Fernandez-Perez ER, Solomon J, Sprunger D, et al. Sarcoidosis-related mortality in the United States from 1988 to 2007. *Am J Respir Crit Care Med*. 2011;183:1524–30.

179. Falasinnu T, Rossides M, Chaichian Y, Simard JF. Do death certificates underestimate the burden of rare diseases? The example of systemic lupus erythematosus mortality, Sweden, 2001-2013. *Public Health Rep.* 2018;133:481–8.
180. Tukey MH, Berman JS, Boggs DA, White LF, Rosenberg L, Cozier YC. Mortality among African American women with sarcoidosis: data from the Black Women’s Health Study. *Sarcoidosis Vasc Diffus Lung Dis.* 2013;30:128–33.
181. Pick HJ, Bolton CE, Lim WS, McKeever TM. Patient-reported outcome measures in the recovery of adults hospitalised with community-acquired pneumonia: a systematic review. *Eur Respir J.* 2019;53:1802165.
182. Tong S, Amand C, Kieffer A, Kyaw MH. Trends in healthcare utilization and costs associated with pneumonia in the United States during 2008–2014. *BMC Health Serv Res.* 2018;18:715.
183. Dureault A, Chapelon C, Biard L, Domont F, Savey L, Bodaghi B, et al. Severe infections in sarcoidosis: incidence, predictors and long-term outcome in a cohort of 585 patients. *Med.* 2017;96:e8846.
184. Uzunhan Y, Nunes H, Jeny F, Lacroix M, Brun S, Brillet P-Y, et al. Chronic pulmonary aspergillosis complicating sarcoidosis. *Eur Respir J.* 2017;49:1602396.
185. Ungprasert P, Crowson CS, Matteson EL. Sarcoidosis increases risk of hospitalized infection. A population-based study, 1976–2013. *Ann Am Thorac Soc.* 2017;14:676–81.
186. Chovatiya R, Silverberg JI. Association of herpes zoster and chronic inflammatory skin disease in US inpatients. *J Am Acad Dermatol.* 2020;[In press].
187. Bloom MW, Greenberg B, Jaarsma T, Januzzi JL, Lam CSP, Maggioni AP, et al. Heart failure with reduced ejection fraction. *Nat Rev Dis Prim.* 2017;3:17058.
188. Bargagli E, Rosi E, Pistolesi M, Lavorini F, Voltolini L, Rottoli P. Increased Risk of Atherosclerosis in Patients with Sarcoidosis. *Pathobiology.* 2017;84:258–63.
189. Zöller B, Li X, Sundquist J, Sundquist K. Risk of subsequent coronary heart disease in patients hospitalized for immune-mediated diseases: a nationwide follow-up study from Sweden. *PLoS One.* 2012;7:e33442.
190. Ungprasert P, Crowson CS, Matteson EL. Risk of cardiovascular disease among patients with sarcoidosis: a population-based retrospective cohort study, 1976–2013. *Eur Respir J.* 2017;49:1601290.
191. Clarson LE, Bajpai R, Whittle R, Belcher J, Abdul Sultan A, Kwok CS, et al. Interstitial lung disease is a risk factor for ischaemic heart disease and myocardial infarction. *Heart.* 2020;106:916–22.
192. Yafasova A, Fosbøl EL, Schou M, Gustafsson F, Rossing K, Bundgaard H, et al. Long-Term Adverse Cardiac Outcomes in Patients With Sarcoidosis. *J Am Coll Cardiol.* 2020;76:767–77.
193. Bonifazi M, Bravi F, Gasparini S, La Vecchia C, Gabrielli A, Wells AU, et al. Sarcoidosis and cancer risk: systematic review and meta-analysis of observational studies. *Chest.* 2015;147:778–91.

194. Ungprasert P, Crowson CS, Matteson EL. Risk of malignancy among patients with sarcoidosis: a population-based cohort study. *Arthritis Care Res (Hoboken)*. 2017;69:46–50.
195. Søgaaard KK, Sværke C, Thomsen RW, Nørgaard M. Sarcoidosis and subsequent cancer risk: a Danish nationwide cohort study. *Eur Respir J*. 2015;45:269–72.
196. Drent M, Wirnsberger RM, Breteler MH, Kock LM, de Vries J, Wouters EF. Quality of life and depressive symptoms in patients suffering from sarcoidosis. *Sarcoidosis, Vasc Diffus Lung Dis*. 1998;15:59–66.
197. Hinz A, Brähler E, Möde R, Wirtz H, Bosse-Henck A. Anxiety and depression in sarcoidosis: the influence of age, gender, affected organs, concomitant diseases and dyspnea. *Sarcoidosis, Vasc Diffus Lung Dis*. 2012;29:139–46.
198. Goracci A, Fagiolini A, Martinucci M, Calossi S, Rossi S, Santomauro T, et al. Quality of life, anxiety and depression in sarcoidosis. *Gen Hosp Psychiatry*. 2008;30:441–5.
199. Balaji AL, Abhishekh HA, Kumar NC. Depression in sarcoidosis patients of a tertiary care general hospital. *Aust New Zeal J Psychiatry*. 2012;46:1199–200.
200. Vandenbroucke JP, Pearce N. Case-control studies: basic concepts. *Int J Epidemiol*. 2012;41:1480–9.
201. Rothman KJ, Lash TL, Greenland S. *Modern epidemiology*. 3rd ed. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
202. Hernán MA, Robins JM. Target trial emulation. *Causal Inference What If*. Boca Radon, FL: Chapman & Hall/CRC; 2020. p. 277–87.
203. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 2016;79:70–5.
204. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183:758–64.
205. Tenesa A, Haley CS. The heritability of human disease: estimation, uses and abuses. *Nat Rev Genet*. 2013;14:139–49.
206. Frisell T, Saevarsdottir S, Askling J. Family history of rheumatoid arthritis: an old concept with new developments. *Nat Rev Rheumatol*. 2016;12:335–43.
207. Hernán MA, Robins JM. *Causal Inference: What If*. 1st ed. Boca Radon, FL: Chapman & Hall/CRC; 2020.
208. Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. *CHANCE*. 2019;32:42–9.
209. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York City, NY: Springer; 2009.
210. Lash TL, Fox MP, Maclehose RF, Maldonado G, Mccandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014;43:1969–85.
211. Pearce N. Analysis of matched case-control studies. *BMJ*. 2016;352:i969.

212. Hosmer D, Lemeshow S, Sturdivant R. Applied Logistic Regression. 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc.; 2013.
213. Hosmer DW, Lemeshow S, May S. Applied Survival Analysis: Regression Modeling of Time-to-Event Data. 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc.; 2008.
214. Royston P, Lambert PC. Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model. College Station, TX: Stata Press; 2011.
215. Amorim LDAF, Cai J. Modelling recurrent events: A tutorial for analysis in epidemiology. *Int J Epidemiol*. 2015;44:324–33.
216. van der Laan MJ, Rose S. Targeted Learning in Data Science. Cham: Springer International Publishing; 2018.
217. Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. *Am J Epidemiol*. 2017;185:65–73.
218. Luque-Fernandez MA, Schomaker M, Rachet B, Schnitzer ME. Targeted maximum likelihood estimation for a binary treatment: a tutorial. *Stat Med*. 2018;37:2530–46.
219. Ceder S, Rossides M, Kullberg S, Eklund A, Grunewald J, Arkema E V. Positive predictive value of sarcoidosis identified in an administrative healthcare registry: a validation study. *Epidemiology*. 2021;[In Press].
220. Cederkvist L, Holst KK, Andersen KK, Glidden D V, Frederiksen K, Kjaer SK, et al. Incorporation of the time aspect into the liability-threshold model for case-control-family data. *Stat Med*. 2017;36:1599–618.
221. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009;20:512–22.
222. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis*. 1987;40:373–83.
223. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed*. 2004;75:45–9.
224. Pagnoux C, Mahr A, Hamidou MA, Boffa JJ, Ruivard M, Ducroix JP, et al. Azathioprine or methotrexate maintenance for ANCA-associated vasculitis. *N Engl J Med*. 2008;359:2790–803.
225. Schram ME, Roekevisch E, Leeftang MMG, Bos JD, Schmitt J, Spuls PI. A randomized trial of methotrexate versus azathioprine for severe atopic eczema. *J Allergy Clin Immunol*. 2011;128:353–9.
226. Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *Am J Epidemiol*. 2004;159:702–6.
227. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med*. 2017;167:268.
228. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim J-L, Reuterwall C, et al. External review and validation of the Swedish National Inpatient Register. *BMC Public Health*. 2011;11:450.

229. Sjölander A, Vansteelandt S. Doubly robust estimation of attributable fractions in survival analysis. *Stat Methods Med Res.* 2017;26:948–69.
230. Statistiska Centralbyrån. Befolkningsstatistik [Population statistics] [Internet]. 2018 [cited Oct 1, 2020]. Available from: <https://www.scb.se>
231. Rossides M, Grunewald J, Eklund A, Kullberg S, Askling J, Arkema E V. Correspondence for “Clinical epidemiology of familial sarcoidosis: A systematic literature review.” *Respir Med.* 2019;160:105696.
232. Zhou Y, Hu Y, Li H. Role of *Propionibacterium acnes* in sarcoidosis: a meta-analysis. *Sarcoidosis, Vasc Diffus Lung Dis.* 2013;30:262–7.
233. Fang C, Huang H, Xu Z. Immunological Evidence for the Role of *Mycobacteria* in Sarcoidosis: A Meta-Analysis. *PLoS One.* 2016;11:e0154716.
234. Arkema E V, Goldstein BL, Robinson W, Sokolove J, Wagner CA, Malspeis S, et al. Anti-citrullinated peptide autoantibodies, human leukocyte antigen shared epitope and risk of future rheumatoid arthritis: a nested case–control study. *Arthritis Res Ther.* 2013;15:R159.
235. Mooney J, Chang E, Lalla D, Papoyan E, Raimundo K, Reddy SR, et al. Potential delays in diagnosis of idiopathic pulmonary fibrosis in Medicare beneficiaries. *Ann Am Thorac Soc.* 2019;16:393–6.
236. Eriksson C, Kokkonen H, Johansson M, Hallmans G, Wadell G, Rantapää-Dahlqvist S. Autoantibodies predate the onset of systemic lupus erythematosus in northern Sweden. *Arthritis Res Ther.* 2011;13:R30.
237. Rodrigues MM, Coletta ENAM, Ferreira RG, Pereira CA de C. Delayed diagnosis of sarcoidosis is common in Brazil. *J Bras Pneumol.* 2013;39:539–46.
238. Oswald-Richter K, Drake W. The etiologic role of infectious antigens in sarcoidosis pathogenesis. *Semin Respir Crit Care Med.* 2010;31:375–9.
239. Kirkil G, Lower EE, Baughman RP. Predictors of Mortality in Pulmonary Sarcoidosis. *Chest.* 2018;153:105–13.
240. Jeny F, Uzunhan Y, Lacroix M, Gille T, Brillet P-Y, Nardi A, et al. Predictors of mortality in fibrosing pulmonary sarcoidosis. *Respir Med.* 2020;169:105997.
241. Kearney GD, Obi ON, Maddipati V, Mohan A, Malur A, Carter JC, et al. Sarcoidosis deaths in the United States: 1999–2016. *Respir Med.* 2019;149:30–5.
242. Feldman CH, Hiraki LT, Winkelmayer WC, Marty FM, Franklin JM, Kim SC, et al. Serious infections among adult Medicaid beneficiaries with systemic lupus erythematosus and lupus nephritis. *Arthritis Rheumatol.* 2015;67:1577–85.
243. Greulich S, Deluigi CC, Gloekler S, Wahl A, Zürn C, Kramer U, et al. CMR Imaging Predicts Death and Other Adverse Events in Suspected Cardiac Sarcoidosis. *JACC Cardiovasc Imaging.* 2013;6:501–11.
244. Shlobin OA, Kouranos V, Barnett SD, Alhamad EH, Culver DA, Barney J, et al. Physiological predictors of survival in patients with sarcoidosis-associated pulmonary hypertension: results from an international registry. *Eur Respir J.* 2020;55:1901747.

245. Tiosano S, Versini M, Dar Antaki L, Spitzer L, Yavne Y, Watad A, et al. The long-term prognostic significance of sarcoidosis-associated pulmonary hypertension – A cohort study. *Clin Immunol*. 2019;199:57–61.
246. Patterson KC, Strek ME. Pulmonary Fibrosis in Sarcoidosis. Clinical Features and Outcomes. *Ann Am Thorac Soc*. 2013;10:362–70.
247. Helby J, Nordestgaard BG, Benfield T, Bojesen SE. Asthma, other atopic conditions and risk of infections in 105 519 general population never and ever smokers. *J Intern Med*. 2017;282:254–67.
248. Widdifield J, Bernatsky S, Paterson JM, Gunraj N, Thorne JC, Pope J, et al. Serious infections in a population-based cohort of 86,039 seniors with rheumatoid arthritis. *Arthritis Care Res (Hoboken)*. 2013;65:353–61.
249. Listing J, Gerhold K, Zink A. The risk of infections associated with rheumatoid arthritis, with its comorbidity and treatment. *Rheumatology*. 2013;52:53–61.
250. Feldman CH, Marty FM, Winkelmayer WC, Guan H, Franklin JM, Solomon DH, et al. Comparative rates of serious infections among patients with systemic lupus erythematosus receiving immunosuppressive medications. *Arthritis Rheumatol*. 2017;69:387–97.
251. Dixon WG, Abrahamowicz M, Beauchamp M-E, Ray DW, Bernatsky S, Suissa S, et al. Immediate and delayed impact of oral glucocorticoid therapy on risk of serious infection in older patients with rheumatoid arthritis: a nested case-control analysis. *Ann Rheum Dis*. 2012;71:1128–33.
252. Dixon WG, Suissa S, Hudson M. The association between systemic glucocorticoid therapy and the risk of infection in patients with rheumatoid arthritis: systematic review and meta-analyses. *Arthritis Res Ther*. 2011;13:R139.
253. Desai RJ, Bateman BT, Huybrechts KF, Paterno E, Hernandez-Diaz S, Park Y, et al. Risk of serious infections associated with use of immunosuppressive agents in pregnant women with autoimmune inflammatory conditions: cohort study. *BMJ*. 2017;356:j895.
254. George MD, Baker JF, Winthrop K, Hsu JY, Wu Q, Chen L, et al. Risk for Serious Infection With Low-Dose Glucocorticoids in Patients With Rheumatoid Arthritis. *Ann Intern Med*. 2020;173:870–8.
255. Jamilloux Y, Valeyre D, Lortholary O, Bernard C, Kerever S, Lelievre L, et al. The spectrum of opportunistic diseases complicating sarcoidosis. *Autoimmun Rev*. 2015;14:64–74.
256. Arnold MH, O’Callaghan J, McCredie M, Beller EM, Kelly DE, Brooks PM. Comparative controlled trial of low-dose weekly methotrexate versus azathioprine in rheumatoid arthritis: 3-year prospective study. *Rheumatology*. 1990;29:120–5.
257. Hamdy H, McKendry RJR, Mierins E, Liver JA. Low-dose methotrexate compared with azathioprine in the treatment of rheumatoid arthritis: a twenty-four-week controlled clinical trial. *Arthritis Rheum*. 1987;30:361–8.
258. Jeurissen MEC, Agnes AM, Van De Putte LBA, Doesburg WH, Mulder J, Rasker JJ, et al. Methotrexate versus azathioprine in the treatment of rheumatoid arthritis. A forty-eight-week randomized, double-blind trial. *Arthritis Rheum*. 1991;34:961–72.

259. Bernatsky S, Hudson M, Suissa S. Anti-rheumatic drug use and risk of serious infections in rheumatoid arthritis. *Rheumatology*. 2007;46:1157–60.
260. Solomon DH, Glynn RJ, Karlson EW, Lu F, Corrigan C, Colls J, et al. Adverse Effects of Low-Dose Methotrexate. *Ann Intern Med*. 2020;172:369.
261. Björk J, Andersson T, Ahlbom A. Commonly used estimates of the genetic contribution to disease are subject to the same fallacies as bad luck estimates. *Eur J Epidemiol*. 2019;34:987–92.
262. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol*. 1985;122:904–14.
263. Månsson R, Joffe MM, Sun W, Hennessy S. On the estimation and use of propensity scores in case-control and case-cohort studies. *Am J Epidemiol*. 2007;166:332–9.
264. Guertin JR, Rahme E, Dormuth CR, LeLorier J. Head to head comparison of the propensity score and the high-dimensional propensity score matching methods. *BMC Med Res Methodol*. 2016;16:22.
265. Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. *Am J Epidemiol*. 1988;128:1185–97.
266. Mansournia MA, Altman DG. Population attributable fraction. *BMJ*. 2018;360:k757.
267. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *Am J Public Health*. 1998;88:15–9.
268. Kahlmann V, Janssen Bonás M, Moor CC, van Moorsel CHM, Kool M, Kraaijvanger R, et al. Design of a randomized controlled trial to evaluate effectiveness of methotrexate versus prednisone as first-line treatment for pulmonary sarcoidosis: the PREDMETH study. *BMC Pulm Med*. 2020;20:271.